Principles of Remote Sensing

An introductory textbook

Editors

Klaus Tempfli
Norman Kerle
Gerrit C. Huurneman
Lucas L. F. Janssen

Authors of Text

Wim H. Bakker
Wim Feringa
Ambro S. M. Gieske
Ben G. H. Gorte
Karl A. Grabmaier
Chris A. Hecker
John A. Horn
Gerrit C. Huurneman
Lucas L. F. Janssen
Norman Kerle
Freek D. van der Meer
Gabriel N. Parodi
Christine Pohl
Colin V. Reeves
Frank J. van Ruitenbeek
Ernst M. Schetselaar
Klaus Tempfli
Michael J. C. Weir
Eduard Westinga
Tsehaie Woldai

Cover illustration:

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 Geo-Information Science and Earth Observation (ITC),
Hengelosestraat 99,
P.O. Box 6,
7500 AA Enschede, The Netherlands

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Principles of Remote Sensing
Klaus Tempfli, Norman Kerle, Gerrit C. Huurneman and Lucas L. F. Janssen (eds.)

(ITC Educational Textbook Series; 2)
Fourth edition

ISBN 978–90–6164–270–1 ITC, Enschede, The Netherlands
ISSN 1567–5777 ITC Educational Textbook Series

Contents

List of Figures

List of Tables

Preface

Principles of Remote Sensing is written to be primarily used as textbook for the introductory module on Earth Observation of all ITC degree courses and for several short diploma or certificate courses. Together with the complementary book Principles of Geographic Information Systems it can also serve as a basic reference on geoinformatics for a wide disciplinary user range from urban and regional planning, civil engineering and geo-engineering, management of natural resources, environmental protection, disaster management and defense, water management and earth resources exploration to land administration and governance. The book presents the basic principles of geospatial data acquisition limiting the scope to the intersection of earth observation and remote sensing. The design of the book was governed by the aim to cover the interests of a very wide user spectrum and the restriction that a novice student of geoinformatics can learn the selected basics in three weeks. The structure of the book follows the classical set-up of first reviewing the necessary physics before discussing sensor technology and introducing common image processing techniques. The title Remote Sensing has been kept for historical reasons; the book is now in its forth edition.

You may wonder why ITC has been producing its own introductory textbook while there are already many books on the subject available on the market. PoRS 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. Being used at the beginning of the courses it tries to stimulate conceptual and abstract thinking without overloading it with formulae. The book relates to the typical ITC application disciplines and is, therefore, not limited to traditional sectorial remote sensing; it includes photogrammetric subjects and an introduction into techniques of data acquisition for subsurface characteristics. Finally, compared to other introductory books, which often focus on the technique, PoRS 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 hardcopy 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 testing. The book comprises a glossary, an index and a bibliography. The electronic document (PDF format) enables fast navigation and quick referencing.

Acknowledgements

Lucas Janssen was the editor of the first edition of *Principles of Remote Sensing*. In 2000 he wrote in his preface to the book:

"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 hardcopy 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."

Preface to the fourth edition

Gerrit Huurneman was the editor of the second edition of PoRS. Gerrit added a new chapter on atmospheric aspects, and revised and updated the entire text, in particular the information on multispectral scanners, the chapter that becomes outdated most quickly. Three years later it was Norman Kerle, who took up the editorial work for the third edition. Norman updated again the chapter on scanners and accomplished a complete overhaul of the chapter on radiometric correction. Moreover, he had new sections and chapters added, which accounted for new developments and could serve as reference material for more specific concepts and methods beyond sole core matter. The chapters on laser scanning, thermal remote sensing, and the one on imaging spectrometry were entirely new.

The third edition served as textbook for five years while ICT continued to rapidly develop. Taking into account the suggestions of many ITC colleagues I decided to regroup several chapters to better "map" technological innovation to ITC's courses. Updating, focusing and homogenizing were the aims of the revision for the 4^{th} edition, not further extending the content of the book. The previous chapters on scanners and digital aerial cameras are now merged and specifications of current sensors have migrated to an Appendix. The new chapter is more focused on sensor system characteristics. I have added the basic principles of sensing to the physics chapter, which as a whole has become more quantitative. The previous chapters on radiometric correction and image enhancement have been rearranged to a core chapter on visualization and radiometric corrections. The chapter on visual image interpretation has become shorter, now describing

the principles of interpretation and the general approach rather than elaborating on application examples.

The chapters not belonging to the core subjects of all ITC courses are grouped in the *Capita Selecta* Section in the second part of the book (*ie*, Chapters 9 to 14). There is only on exception, *ie*, the elaboration on 'image fusion' in Section 5.6, which is also a subject of a course specific choice.

I have made a major effort to homogenize the terminology and significantly extend the glossary. You find there not only the definitions of disciplinary terms but also – where deemed appropriate – a sharpened or more elaborate description of a notion than in the text body of the book. New is the list of abbreviations and list of formulae. If you should doubt the English here and there, take it as the Europeanized version of English. Finally, many of the corrigenda have been fixed for the softcopy version of the book.

I would like to thank my colleagues, who made many valuable suggestions for updating the book, in particular Markus Gerke, Chris Hecker, Norman Kerle, Monika Kuffer, Georges Vosselman, Michael Weir, and Eduard Westinga – the latter I want to also thank for giving shape to the new chapter on interpretation.

Above all I want to thank Coco Rulinda for doing all the latexing: without her generous and dedicated support revising the book would have been impossible.

Lastly I would like to thank my wife, Gerti, for tolerating me getting carried

away with the book on a couple of evenings and weekends.

Klaus Tempfli, Enschede, September 28, 2008

Common core

Chapter 1

Introduction to earth observation by remote sensing

1.1 Geospatial data acquisition

Are we making sufficient progress on social well-being, sustainable economic development, and environmental protection? Do we have the right information available at the right time at the right place to take adequate decisions? Instrumental to achieving such progress is the availability of geospatial data, the more so the more drastically we exploit the planet Earth. Fortunately, we can notice that the awareness increases worldwide of the importance of having access to reliable, detailed, timely and affordable geospatial data. Equally important, the technology of acquiring and providing geospatial data quickly develops stimulated by the rapid advances of information and communication technology.

This book is about basic principles of geospatial data acquisition. Geospatial data acquisition is central to earth observation. *Earth observation* is gathering of information about physical, chemical, biological, geometrical properties of our planet; it helps us to assess the status and monitor changes of the natural and cultural environment. Thus, mapping, monitoring, and also forecasting are the uses of earth observation. Earth observation gives us geospatial data. Geospatial data acquisition can be taken as a starting point in our development cycle of observing - analyzing - designing or planning - constructing or developing - observing, etc. Geospatial data are very valuable, so we rather not consider their acquisition as the first stage of a linear system - as we know it from materials economy - which starts with acquisition and ends via production and consumption at disposal.

Earth observation

The following examples illustrate the diverse need for geospatial data:

- A land administrator should have an up-to-date record of property boundaries. Property boundaries usually coincide with observable terrain features, which s/he needs to have surveyed.

- A civil engineer has to design a highway. One of the criteria for finding an optimum alignment of the new road is to balance cut and fill volumes in road construction. Thus, s/he needs information about the shape of the ground surface. Calculation of actual transportation costs of material can be based on re-surveying terrain relief after construction.

- An urban planner may want to identify areas of informal settlement. The different types of houses and their configuration need to be determined. The municipality may furnish infrastructural improvements based on a development plan for the identified areas. The urban planner will have to monitor the impact of the provisions before proceeding to further planning.

- An agronomist is interested in forecasting the overall agricultural production of a large area. S/he needs the size of fields per crop and data on biomass production to estimate the yield. Observing soil properties and monitoring degradation will improve forecasts.

- An environment analyst is worried about pollutants of waste disposal sites. S/he has to detect the dump composition and determine volumes.

- A climatologist would like to understand the El Niño phenomenon. To this end s/he needs data on spatial patterns of sea surface temperature at different times, data on sea levels and sea currents, the direction and

velocity of surface winds, information about the processes of interaction between ocean and land surfaces, etc.

These few examples indicate already that our earth observation interest is on objects, scenes, and phenomena, which are different in spatial extent. Moreover, we are interested in different aspects of objects, different levels of detail, and all that at various times and different repeatability rates. Accordingly we use a wide range of different methods and techniques of earth observation. Also obvious from the above examples may be the fact that solving higher order problems, such as sustainable economic development, requires more data than those we can obtain by earth observation. To satisfy the information requirements of geo-scientists, engineers, land and water managers we conduct interviews (*eg*, in adjudication for land administration), take field samples and analyze them in laboratories, observe using in situ sensors (*eg*, a tide gauge), do land surveying, make aerial surveys using cameras and scanners, or we take a more synoptic view by an instrument on a space shuttle, an orbiting space station, a reconnaissance satellite, or an earth resources satellite. We use meteorological satellites for forecasting weather; Meteosat looks at our planet from really remote outer space at some 36,000 km above mean sea level.

The following should help to delineate the subject area of this book and place some terms, which you will come across in literature. You may find them being used by some authors indifferently or by others with a specific but geo-discipline dependent flavour. In this book we try to give them a specific but not discipline dependent meaning and use them consistently.

Remote sensing (RS) has been defined in many different ways. A sufficient definition for this book is: *remote sensing* is the art, science, and technology of observing an object, scene, or phenomenon by instrument-based techniques. 'Remote' because observation is done at a distance without physical contact with the object of interest. We can either use detection and real-time display devices or recording devices of energy, which is emitted or reflected from an object or a scene. The energy can be light or another form of electromagnetic radiation, force fields, or acoustic energy. An example of a remote sensor is a conventional camera. Light reflected from an object passes through the lens and the light-sensitive film detects it. At the moment of exposure a latent image is recorded. Developing and fixing the film in the photo lab generates a definite record, the photograph. This image is then subject to interpretation. Today most remote sensors are electronic devices. The data recorded by such sensors, *eg*, a scanner detecting thermal emission (heat), used to be converted to images - for visual interpretation. Accordingly we still refer to the record produced by an electronic sensor as (remote sensing) image.

Remote sensing

RS defined as above is applied in many fields, including architecture, archeology, medicine, industrial quality control, robotics, extraterrestrial mapping, etc. The interest domain for this book is, however, earth observation and even more specifically earth observation from airborne or spaceborne platforms. Earth observation does not only rely on RS but also on sensors that allow us to make in situ observations. The principles of sensing in physical contact with an object are beyond the scope of this book. Limiting the spatial interest of RS to objects that can (and are relevant to) be located on the surface of the Earth - using a geodetically defined coordinate system - we use the term *geospatial data acquisition* (GDA). The outcome of GDA is not simply an image as obtained by

Geospatial data acquisition

converting sensor recordings, but adequately processed or interpreted data from the sensor. Processing/interpretation and validating require knowledge about the sensing process and yields data readily suited for analysis, *eg*, in a GIS. Typical products derived from geospatial data are orthophoto maps, "satellite image maps", topographic maps, thematic maps such as land use maps, land use change statistics, etc.

Historical perspective

Placing geospatial data acquisition in the historical perspective we can take Surveying and Mapping as starting point. About a century ago Photogrammetry evolved as a sub-discipline of surveying and mapping, offering an extension to the traditional ground-based methods. Photogrammetry and other aerial surveying techniques could replace to a large extent observing directly in the field by measuring terrain features indirectly, on images in the office. The next technological extension was Remote Sensing, which has enabled us to see phenomena our eyes cannot see. We can detect, *eg*, thermal emission and display it such that we can analyze it with our eye-brain system. The professional organization International Society of Photogrammetry was renamed to International Society of Photogrammetry and Remote Sensing (ISPRS) in 1980. Technological development continued; now we have a journal of Earth Observation and Remote Sensing. The ITC Journal, which reported on research in photogrammetry, photo interpretation, and cartography became the International Journal of Applied Earth Observation and Geoinformation.

1.2 Remote sensing for earth observation

Why remote sensing, what is different from other ways of earth observation? This section will generically answer the question using the example of El Niño.

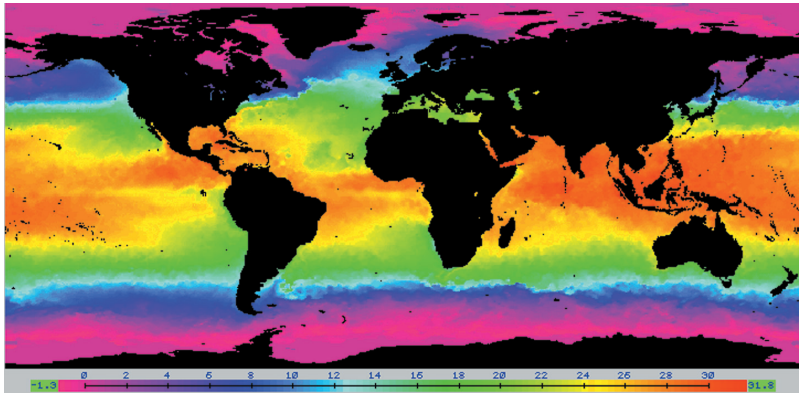


Figure 1.1: Sea surface temperature as determined from NOAA-AVHRR data. Courtesy of NOAA.

Global warming and the rapid change of climate have an enormous impact on our well-being. In early 1998 we observed particularly abnormal weather in many parts of the world. There was very heavy rain in otherwise dry areas, causing landslides in the Andes. There was drought and huge forest fires in Indonesia. The devastating weather coincided with a strong rise of sea water temperature in the Eastern Pacific Ocean. Fishermen of Peru have named the phenomenon of water temperature change *El Niño*, because it happens around Christmas, not every year but in intervals of 4 to 9 years. The last El Niño event was in 2006. The 1982 El Niño event caused an estimated economic loss of 8.2 billion $. The most direct impact of the observed cyclic temperature rise is the

RS provides dense data for large areas

interruption of the fishing season in the Eastern Pacific. If we better understood the causes of El Niño and its effects on the climate, we would have a better starting position for taking preventive and mitigation actions. To develop and support a theory, we have to make observations, among others on sea water temperature. How does temperature change in space and time? To find out we could place buoys in the ocean and continuously measure the temperature there. The west to east extent of the Pacific Ocean is roughly 16,000 km. Is the spatial variation of water temperature so small that we get sufficient information by placing buoys with a spacing of 2000 km? Or would we need a network of smaller meshes and possibly a different spacing in the west-east direction and the north-south direction? The thermal scanners on board of the meteorological/environmental NOAA satellites can provide us with data at 1 km spacing. We process the recordings of the scanner - which includes correcting the atmospheric distortions of the thermal emission of the water - to calculate surface temperature and derive sea surface temperature maps (Figure 1.1).

RS benefits from in situ observations

By comparing the temperature measurements at the buoys with the recordings of the scanner we can calibrate the processing and thus obtain accurate temperature values for a network of much higher density than the one of the buoys. This principle of "ground control" applies to most RS methods.

RS observes surface features

A thermal scanner gives us the temperature of the sea surface not the temperature of subsurface currents, while the latter is possible with buoys. In general, the observations by RS relate to a thin layer of the Earth's surface, which may be considered a limitation of RS. Being interested in subsurface features we have to use additional information on how they manifest themselves in surface features.

RS offers high survey repeatability

The European Meteosat provides a new image of the same area every 30 minutes (since 1977). NASA's NOAA satellites have a revisit time of 24 hours.

RS is "multi-channel"

The NOAA satellites do not only have a thermal scanner on board (labelled AVHRR) but also other instruments, which provide us with data on atmospheric temperature, the concentration and distribution of ozone in the stratosphere, etc. To acquire data for studying a complex phenomenon like El Niño we have to rely on sensors on various platforms. NASA's scatterometer aboard the QuikSCAT satellite provides us with information on speed and direction of ocean winds (Figure 1.2). NASA uses OrbView-2 data to study global warming; it can also provide us with ocean biomass estimates (Figure 1.3). We use spaceborne radar and laser altimeters to determine sea level changes (Figure 1.4).

RS is the only way

The Pacific Ocean is known for its unfavourable weather and current 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 a similar sensor. This enables methods for monitoring and change detection. RS has become an important method of earth observation and for many applications it is the only suitable one - since an increasing number of issues are of global concern, such as climate change, environmental degradation, natural disasters, and population growth.

Data on sea surface temperature acquired for studying El Niño could be used

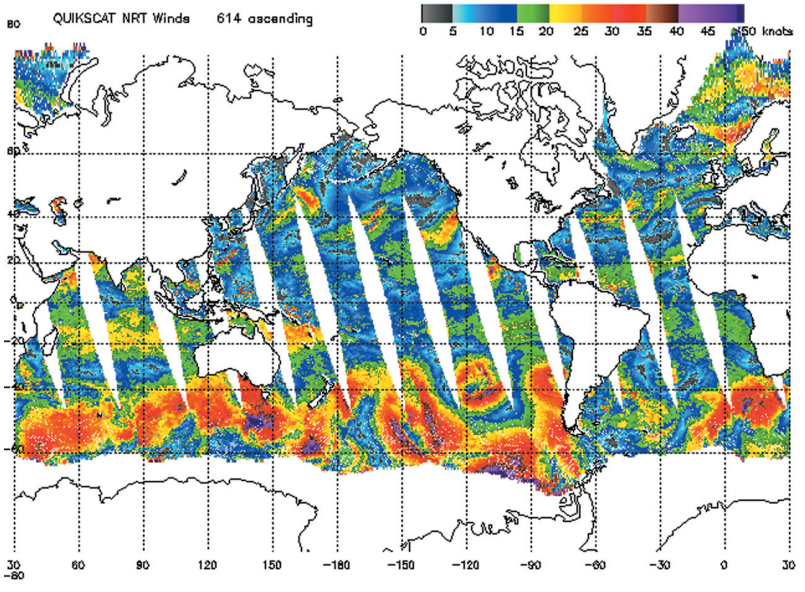


Figure 1.2: Ocean surface wind as determined from scatterometer measurements by QuickSCAT. Courtesy of NOAA.

years later for studying the influences on algae bloom around the Pacific islands. Sea surface temperature maps are not only of interest to researchers; they are used by fishing companies to guide their vessels to promising fishing grounds. Moreover, RS images can also prove useful for studies on phenomena we are not yet (fully) aware of.

RS can serve several purposes

The validity of the statement that RS is cost-effective 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 cost a lot of money. Meteorological satellites have already

RS is cost-effective

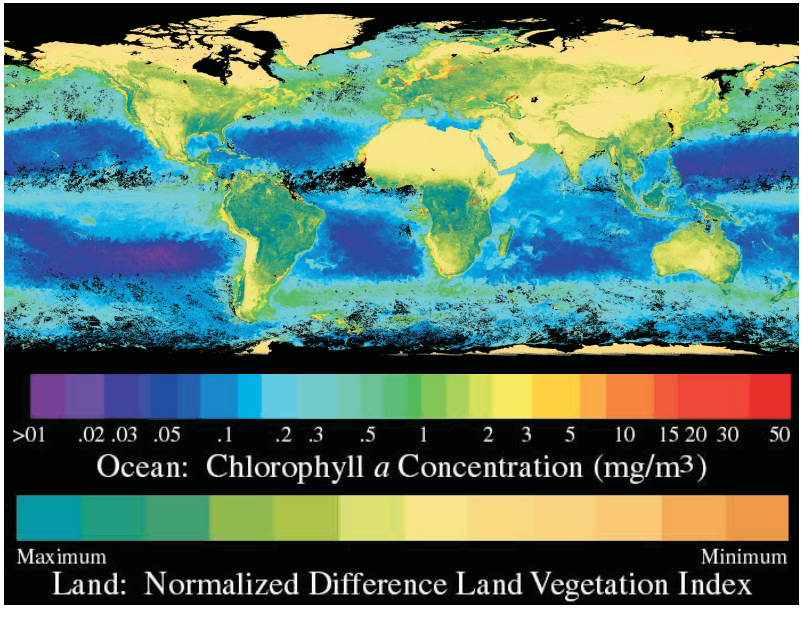


Figure 1.3: Ocean (and land) biomass as determined from OrbView-2 data. Courtesy of Orbimage.

been paid for and the data can be considered "free". Further use of RS then becomes cheap. However, in order to obtain accurate and "in depth" sea water temperature values, we still have to place buoys for calibration purposes and for determining subsurface characteristics.

The above statements on RS characteristics are supported by an example of a scientific problem related to a global phenomenon. However, they are applicable also to other applications. You may analyze yourself whether all the above statements would also apply, *eg*, to the monitoring of (illegal) urban growth

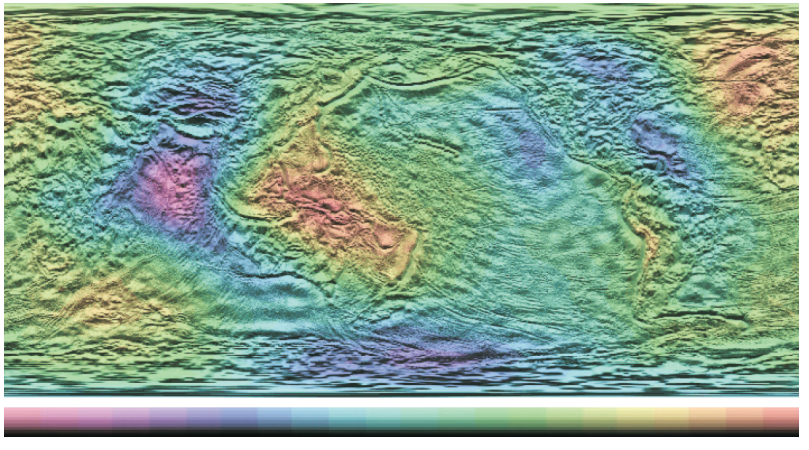


Figure 1.4: Sea surface elevation above mean sea-level as determined from spaceborne radar and laser altimeters. Courtesy of the University of Texas.

around cities using aerial photographs. You will probably 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 meets your own particular information requirements. After studying the textbook *Principles of Remote Sensing* you may expect to be better equipped to consider these issues.

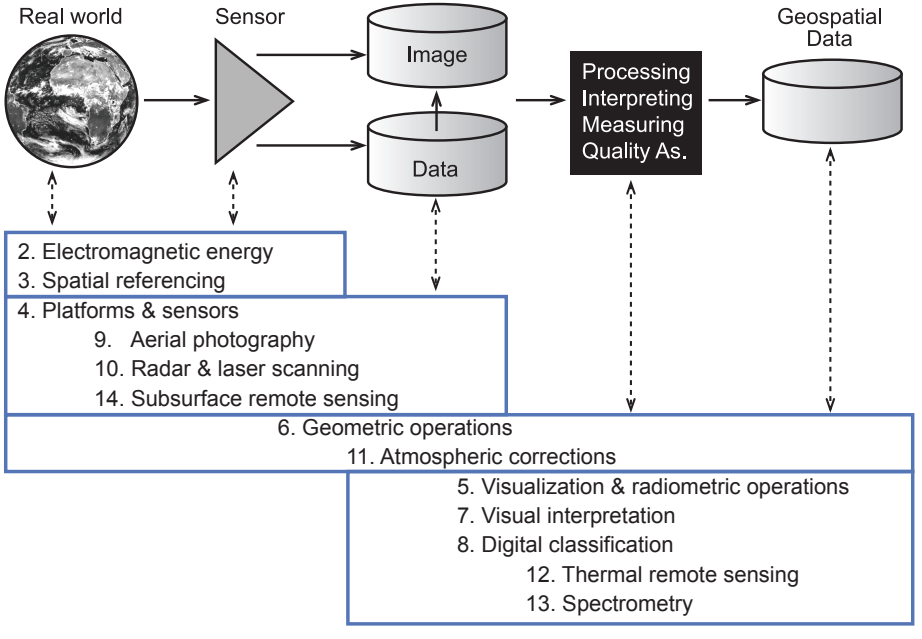


Figure 1.5: Relationship between the chapters of this textbook and the remote sensing process.

1.3 Structure of the textbook

Figure 1.5 relates the chapters of this book to the overall remote sensing process. Chapter 2 reviews the relevant properties of electromagnetic energy, which is reflected or emitted by the Earth's surface and travels through the atmosphere and may be even outer space before reaching the remote sensor. Chapter 2 also introduces the basics of sensing electromagnetic energy. Chapter 3 is very short; it points at determining position on the Earth's surface. Spatial reference systems and spatial positioning are prerequisites for dealing with geometric aspects of earth observation, which is done in Chapter 6. Chapter 4 is on technology, discussing airborne and spaceborne platforms and sensors, which detect and record electromagnetic energy radiated by the Sun or the Earth. Active sensors are sensors which do not depend on an external source of energy - they emit their own and sense what is reflected by the Earth's surface; they are treated in Chapter 10. Chapter 9 is also a special chapter; it reviews aerial photography, which is the basis of the oldest remote sensing method. Sensing alone only provides us with raw data, which need to be processed. We have to apply radiometric processing to obtain data that more readily reveal physical object properties; Chapter 11 introduces methods of 'image restoration' and 'atmospheric corrections'. Chapter 5 deals with different techniques of converting data to actual images and enhancing remote sensing images to facilitate their interpretation. Visual image interpretation and information extraction by algorithm driven classification are introduced in the Chapters 7 and 8. Special Chapters are again the last three (12, 13, 14) on thermal remote sensing, spectrometry, and subsurface remote sensing.

Summary

Many human activities and interests are spatial in nature. We need geospatial data for navigating and planning, for mapping, monitoring, modeling, and decision making. This introductory chapter has exposed the concepts of earth observation, remote sensing (RS), and geospatial data acquisition (GDA). Various examples were given to illustrate the utility of earth observation by remote sensing, indicate the variety of applications and the diversity of data acquisition techniques.

Questions

The following questions can help to study Chapter 1.

1. To what extent is geospatial data acquisition applied by your organization/company?
2. Which methods of in situ observation and which remote sensing methods are applied by your organization/company to acquire geospatial data?
3. Remote sensing data and derived products are available on the internet. Locate three web-based catalogues or archives that comprise remote sensing data.
4. Following the El Niño example, outline how the statements about RS relate to a particular case of your organization (*eg*, land resources, geology, etc).

These are typical exam questions:

1. Explain, or give an example, how in situ 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

GDA challenges us to make choices. On which one of the many sensors should the agronomist of Chapter 1 rely for reliable yield prediction? If it was a sensor producing several images, such as a multispectral scanner, which image or which combination of images to use? How to properly process sensor recordings to increase our chances of correct interpretation? When interpreting a colour image, what causes the sensation 'red'? Instead of writing a thick book of recipes to answer such questions for every application, we better review the physics of RS. Understanding the basics of electromagnetic (EM) energy will help you in making more profound choices and enable you to deal with future sensors.

Figure 2.1: Photograph of the ITC building.

A standard photograph is an image of an object or scene, which very closely resembles direct sensing with our eyes. We see from Figure 2.1 that the roof is

red, the tree is green, the sky is blue. The sensation of colour is caused by EM radiation. Red, green, blue relate to forms of energy which we commonly refer to as light. *Light* is EM radiation that is visible to the human eye. Being interested in earth observation our light source is the Sun. The Sun emits light, the Earth's surface features reflect it, the photosensitive cells (cones and rods) in our eyes detect it. When we look at a photograph, it is the light reflected from it, which lets us observe that the wall of the building is made of bricks. Light is not the only form of energy radiated from the Sun and other bodies. The sensation 'warm' is caused by thermal emission. Sun tanning or our body generating vitamin D is triggered by ultraviolet (UV) radiation.

Section 2.2 explains the basic characteristics of EM energy, its sources, and what we call the EM spectrum. The EM energy which is radiated from the Earth's surface is influenced by the atmosphere. The energy interaction in the atmosphere is described in Section 2.3. Section 2.4 introduces the energy interactions at the Earth's surface. Section 2.5 elaborates on the basic principles of detecting EM energy and generic sensor characteristics.

2.2 Electromagnetic energy

2.2.1 Waves and photons

EM radiation can be modelled in two ways: by waves or by radiant energy bearing particles called photons. The first publications on the wave theory date back to the 17^{th} century. According to the wave theory light travels in a straight line (unless there are outside influences) with energy levels changing in a wave fashion. Light has two oscillating components; the energy constantly changes between electrical energy and magnetic energy. We call it, therefore, electromagnetic energy. The two components interact; an instance of positive electrical energy coincides with a moment of negative magnetic energy (Figure 2.2). The wave behaviour of light is common to all forms of EM radiation. All EM energy travels at the speed of light, which is approximately 300,000 km/s. This is fast, but the distances in space are astronomical. It takes eight minutes for the sunlight to reach the Earth, thus when we see, *eg*, a sunrise, it actually occurred that much earlier. Because of the straight line travel we use the notion of light ray in optics.

EM wave

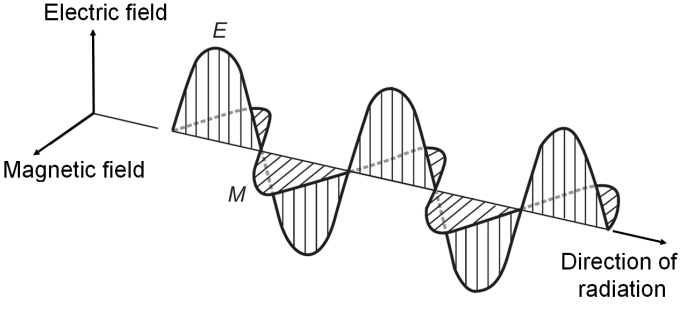


Figure 2.2: The two oscillating components of EM radiation: an electric and a magnetic field.

We can characterize a sine wave by three parameters:

$$e = \alpha sin(\frac{2\pi}{\lambda}x + \varphi). \tag{2.1}$$

The *wavelength*, λ, is the differentiating property of the various types of EM radiation. It is the distance the energy travels from a moment of maximum electrical energy until reaching the maximum again (or any other two corresponding energy states; Figure 2.3). λ is the length of one cycle of the oscillation. It is usually measured in micrometers ($\mu m = 10^{-6}$ m). Blue light is EM radiation with a wavelength of around 0.45 μm. Red light at the other end of the colour spectrum of a rainbow has a wavelength of around 0.65 μm (Figure 2.4). Electromagnetic radiation outside the range 0.38 to 0.76 μm is not visible to the human eye.

Cycle and wavelength

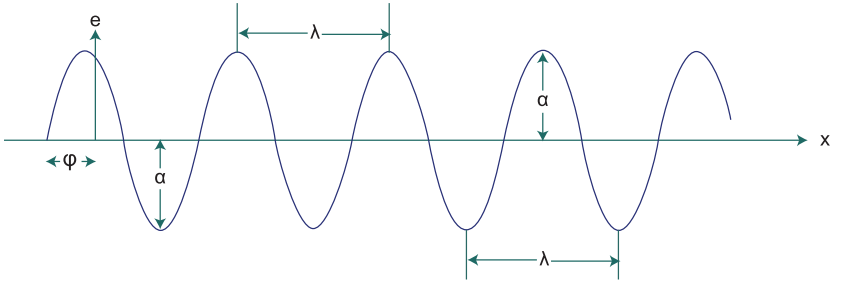


Figure 2.3: Characteristics of a wave, *eg*, an electric wave.

The *amplitude*, α, is the peak value of the wave. The larger the amplitude the higher the energy of the wave (we know this also from water waves). In imaging by active sensors (Chapter 10), the amplitude of the detected signal is used as intensity measure. The *phase*, φ, is an important quantity for precise ranging

Amplitude and phase

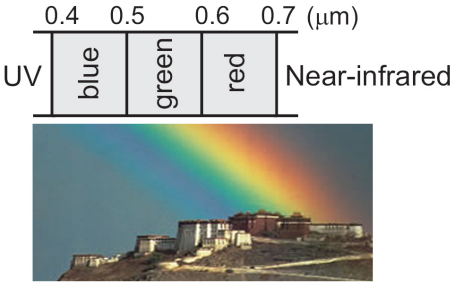


Figure 2.4: The spectrum of light.

(Chapters 3 and 10). The phase expresses the fraction the start point of the wave differs from the origin of counting distance (*ie*, x=0). According to the Formula 2.1 φ can take any value in the range from 0 to 2π.

We call the amount of time needed by an EM wave to complete one cycle the *period* of the wave. The reciprocal of the period is called the *frequency* of the wave. Thus, the frequency is the number of cycles of the wave that occur in one second. We usually measure frequency in hertz (1Hz = 1 cycle per second). The period corresponds to the wavelength of the radiation. Since the speed of light is constant, the relationship between wavelength and frequency is:

Period and frequency

$$c = \lambda \times v. \tag{2.2}$$

Based on Einstein's famous energy formula, the letter c is used commonly as symbol for the speed of light. v is the frequency. Obviously short wavelength implies high frequency, while long wavelength is equivalent to low frequency. Blue light is of higher frequency than red light (Figure 2.5).

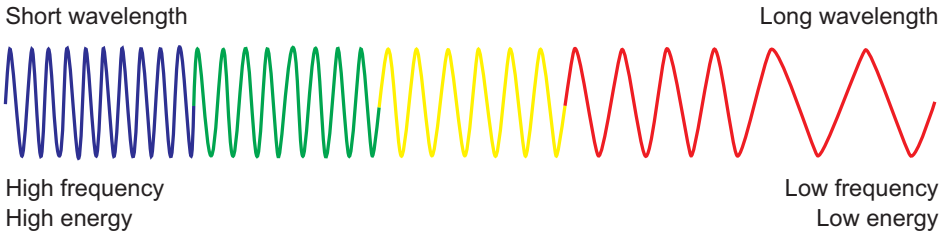


Figure 2.5: Relationship between wavelength, frequency and energy.

We can well explain many EM energy phenomena by the wave theory. However, for some purposes we can better rely on the particle theory, which explains EM energy by photons. We take this approach when quantifying the energy detected by a multispectral sensor (see Section 2.5). The amount of energy held by a photon relating to a specific wavelength is:

Energy of a photon

$$Q = h \times v = h \times \frac{c}{\lambda}, \tag{2.3}$$

where Q is the energy of a photon measured in joule and h is Planck's constant ($6.6262 \cdot 10^{-34}$ jouleseconds).

The energy carried by a single photon of light is just sufficient to excite a single molecule of a photosensitive cell of the human eye, thus contributing to vision. It follows from Equation 2.3 that long wavelength radiation has a low energy level while short wavelength radiation is of high energy. Blue light is more energetic than red light (Figure 2.5). EM radiation beyond violet light is progressively dangerous to our body with increasing frequency. UV radiation can already be harmful to our eyes, so we wear sunglasses to protect them. An important

consequence for RS of the energy relationship expressed by Formula 2.3 is that it is more difficult to detect radiant energy of longer wavelengths than energy of shorter wavelengths.

2.2.2 Sources of EM energy and radiometric units

The Sun is a prime source of EM energy, but it is not the only one. All matter with an absolute temperature above zero emits EM energy because of molecular agitation. Absolute temperature is conventionally measured in kelvins (K) with Celsius-scaled increments. Absolute zero (0 K= $-273.15\,^{\circ}\mathrm{C}$) is the lowest possible temperature where nothing could be colder; at 0 K molecules do not move. The global mean temperature of the Earth's surface is 288 K and the temperature of objects on earth rarely deviates very much from this mean over a finite period. Earth's surface features, therefore, emit EM energy. Solar radiation constantly replenishes the energy that the Earth loses into space. The Sun's temperature is about 6000 K. The Sun emits 44% of its energy as light and 48% as infrared radiation.

The Sun is an approximate black-body and so are stars. A black-body is a theoretical object with assumed extreme properties, which helps us in explaining EM radiation. A black-body absorbs 100% of the radiation that hits it, it does not reflect any; thus, it appears perfectly black. A black-body has the capability of re-emitting all the energy it receives. We say a black-body has the maximum emissivity of 1. A black-body emits energy at every wavelength (Figure 2.6). The energy emitted by a black-body is called black-body radiation. A black-body can have different temperatures. The temperature of the black-body determines the most prominent wavelength of black-body radiation. At room temperature, black-bodies emit predominantly infrared energy. When heating up a black-body beyond 127 K ($1000\,^{\circ}\mathrm{C}$) emission of light becomes dominant, from red, through orange, yellow, and white (at 6000 K) before ending up at blue, be-

yond which the emission includes increasing amounts of ultraviolet radiation. 'White' is special, it is not a colour but a perfect mixture of colour. At 6000 K a black-body emits radiant energy of all visible wavelengths equally. Higher temperature corresponds to a greater contribution of radiation with shorter wavelengths. Figure 2.6 illustrates the physics of what we see when a blacksmith heats a piece of iron, or what we observe when looking at candle. The flame appears light-blue at the outer core; there the flame is the hottest with a temperature of 1670 K. The centre appears orange, with a temperature of 1070 K. More generally, flames may show different colours (depending on the burning material, the temperature of the surrounding, and the amount of oxygen) and accordingly have different temperatures (in the range of $600\,^{\circ}\mathrm{C}$ to $1400\,^{\circ}\mathrm{C}$). Colour tells us about temperature. We can use colour to estimate, *eg*, the temperature of a lava flow - from a safe distance. More general, if we could build sensors, which allow us to detect and quantify EM energy of different wavelengths (also outside the visible range), we could use RS recordings to estimate object temperature. We also notice from the black-body radiation curves (Figure 2.6) that the intensity increases with increasing temperature; the total radiant emittance at a certain temperature is the area under a curve.

Black-body

When quantifying energy we can use different measures. The amount of energy is commonly expressed in joule. We may be interested, however, in the radiant energy per unit time, called the *radiant power*. We measure the power in watt

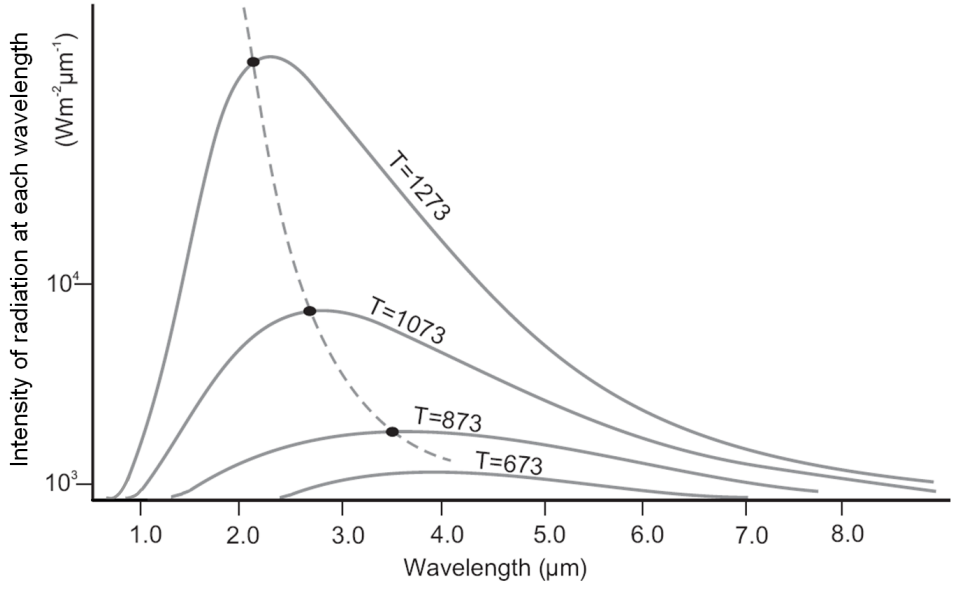


Figure 2.6: Black-body radiation curves (with temperatures, *T*, in K).

(1W = 1 joule per second). *Radiant emittance* (used in the chapter on thermal RS) is the power emitted from a surface; it is measured in watt per square meter (Wm^{-2}). The *spectral radiant emittance* characterizes the radiant emittance per wavelength; it is measured in $Wm^{-2}\mu m^{-1}$ (this is the unit used in Figure 2.6). *Radiance* is another frequently used quantity in RS. It is the radiometric measure, which describes the amount of energy being emitted or reflected from a particular area per unit solid angle and per unit time. Radiance (observed intensity) is usually expressed in $Wm^{-2}sr^{-1}$. *Irradiance* is the amount of incident energy on a surface per unit area and per unit time. Irradiance is usually expressed in Wm^{-2}.

Radiometric units

Real objects can only approximate black-bodies; they can re-emit some 80 to 98% of the received energy. The emitting ability of real material is expressed as dimensionless ratio called emissivity (with values between 0 and 1). The *emissivity* of a material specifies how well a real body emits energy as compared with a black-body. A surface's spectral emissivity depends on several factors, such as temperature, emission angle, and wavelength. Observing the material's emissivity helps us, among others, in modelling global warming; the chapter on thermal RS provides a more detailed discussion on the above outlined concepts.

2.2.3 Electromagnetic spectrum

We call the total range of wavelengths of EM radiation the *EM spectrum*. Figure 2.4 illustrates the spectrum of light, Figure 2.7 the total spectrum of EM radiation. We refer to the different portions of the spectrum by name: gamma rays, X-rays, UV radiation, visible radiation (light), infrared radiation, microwaves, and radio waves. Each of the these named portions represents a range of wavelengths, not one specific wavelength. The EM spectrum is continuous and does not have any clear-cut class boundaries.

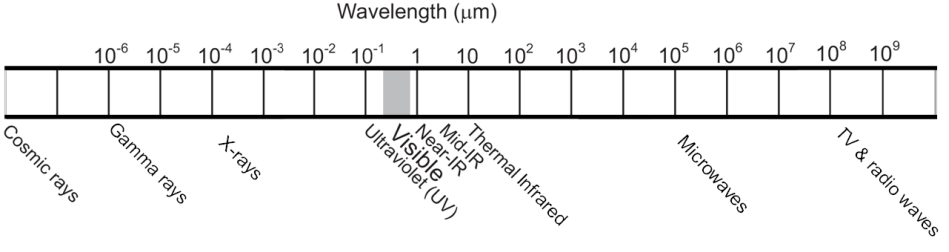


Figure 2.7: The EM spectrum.

Different portion of the spectrum are of different relevance to earth observation both in the type of information that we can gather and the volume of geospatial data acquisition. The majority of GDA is accomplished from sensing in the visible and infrared range. The UV portion covers the shortest wavelengths that are of practical use to earth observation. UV radiation reveals some properties of minerals. At the other end of the useful range for earth observation are the microwaves; they can among others provide information on surface roughness and moisture content of soils.

The 'visible portion' of the spectrum with wavelengths causing colour is only a very small fraction of the entire EM wavelength range. We call objects 'green' when they reflect predominately EM energy of wavelength 0.54 μm. The intensity of solar radiation has its maximum around this wavelength (see Figure 2.10) and the sensitivity of our eyes is peaked at green-yellow. We know that colour effects our emotions and we usually experience green sceneries as pleasant. We use colour to discriminate objects and we can use it to estimate temperature. We also use colour to visualize EM energy we cannot see directly. The image of sea surface temperature (Figure 1.1) shows colours of the total light range from violet to deep red. The chapter on visualization elaborates how we can "produce colour" by adequately "mixing" the three primary colours red, green, and blue.

Light and colour

The radiation beyond red light towards larger wavelengths in the spectrum is referred to as infrared (IR). We can discriminate vegetation types and the stress state of plants by analyzing 'near-infrared' (and 'mid-infrared') radiation - much better then trying to so by colour. For example, deciduous trees reflect more near-infrared (NIR) energy than conifers do, so they show up brighter on photographic film that is sensitive to infrared. Healthy vegetation has a high reflectance in the NIR range, which decreases with increasing damage caused by a plant disease (see also Section 2.4.1). Mid-IR is also referred to as short-wave infrared (SWIR). SWIR sensors are used to monitor surface features at night.

Near-infrared, Short-wave infrared

Infrared radiation with a wavelength longer than 3 μm is termed thermal infrared (TIR), because it causes the sensation of 'heat'. Near-IR and mid-IR do not cause a sensation of something being hot. Thermal emission of the Earth's surface at (300 K) has a peak wavelength of 10 μm (see Figure 12.1). Also a hu-

man body emits 'heat energy', with a maximum at $\lambda \approx 10$ μm. Thermal detectors for humans are, therefore, designed such that they are sensitive to radiation in the wavelength range 7 to 14 μm. NOAA's thermal scanner with its interest in heat issued from the Earth's surface detects thermal IR radiation in the range 3.5 to 12.5 μm (see Appendix A). Surface temperature is needed for studying a variety of environmental problems, useful for analyzing the mineral composition of rocks, the condition of vegetation, etc.

Thermal infrared

2.3 Energy interaction in the atmosphere

Before the Sun's energy reaches the Earth's surface, three RS relevant interactions in the atmosphere happen: absorption, transmission, and scattering. The energy transmitted is then either absorbed by the surface material or reflected. The reflected energy suffers also from scattering and absorption in the atmosphere before reaching the remote sensor (Figure 2.8).

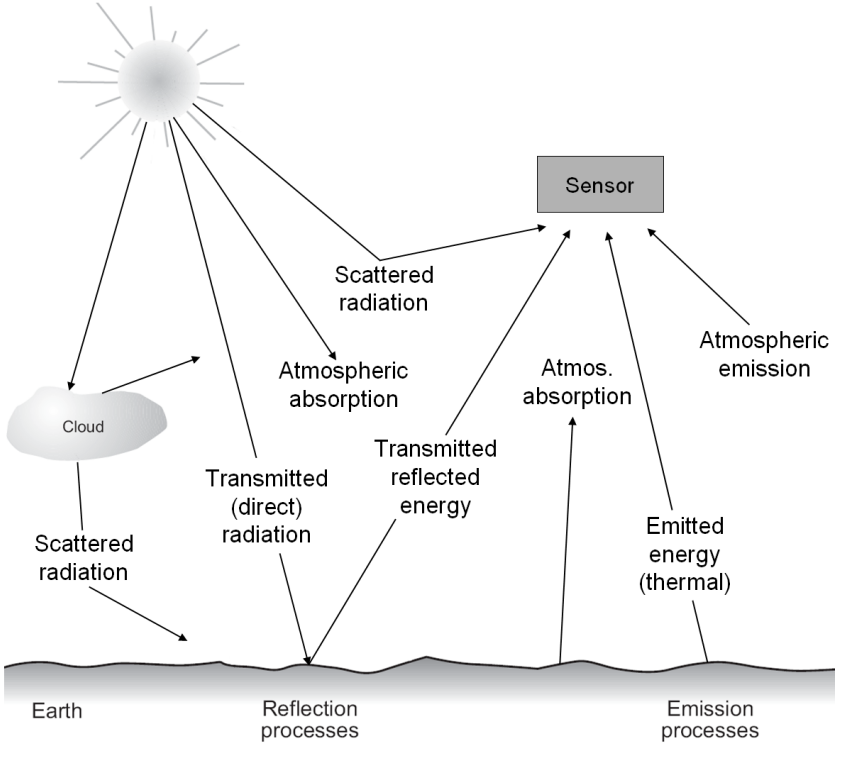


Figure 2.8: Energy interactions in the atmosphere and at the Earth's surface.

2.3.1 Absorption and transmission

Electromagnetic energy traveling 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).

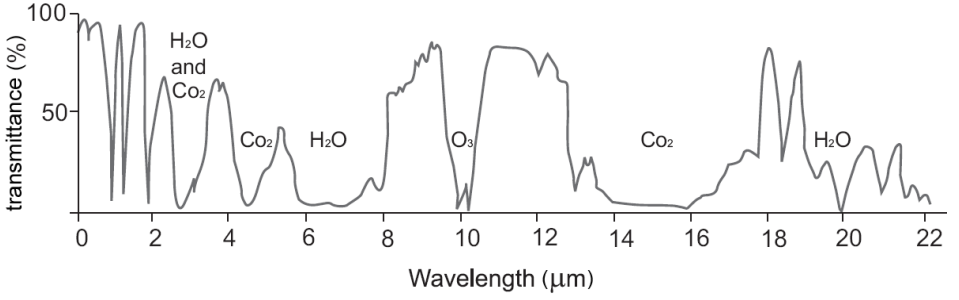


Figure 2.9: Atmospheric transmittance.

Figure 2.9 gives a schematic representation of the atmospheric transmission in the 0 to 22 μm wavelength range. From this figure it may be seen that many of the wavelengths are not useful for remote sensing of the Earth's surface, simply because none of the corresponding energy can penetrate the atmosphere. Only the spectrum portions outside the main absorption ranges of the atmospheric gases can be used for remote sensing. The useful ranges are referred to as the *atmospheric transmission windows* and include:

Atmospheric transmission

- The window from 0.4 to 2 μm. The radiation in this range (visible, NIR, SWIR) is mainly reflected energy. Because this type of radiation follows the laws of optics, remote sensors operating in this range are often referred to as optical ones.

- Three windows in the TIR range, 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 is occurring at longer wavelengths. There is hardly any transmission of energy in the range from 22 μm to 1 mm. The more or less "transparent" range beyond 1 mm is the microwave range.

The solar radiation as observed both with and without the influence of the Earth's atmosphere is shown in Figure 2.10. Solar radiation measured outside the atmosphere resembles the black-body radiation at 6000 K. Measuring solar radiation at the Earth's surface shows that there the spectral distribution of energy radiated from the Sun is very ragged. The relative dips in this curve indicate the absorption by different gases in the atmosphere. We also see from Figure 2.10 that the total intensity in this range (*ie*, the area under the curve) has decreased by the time the solar energy reaches the Earth's surface after having passed through the envelop of gas of increasing density.

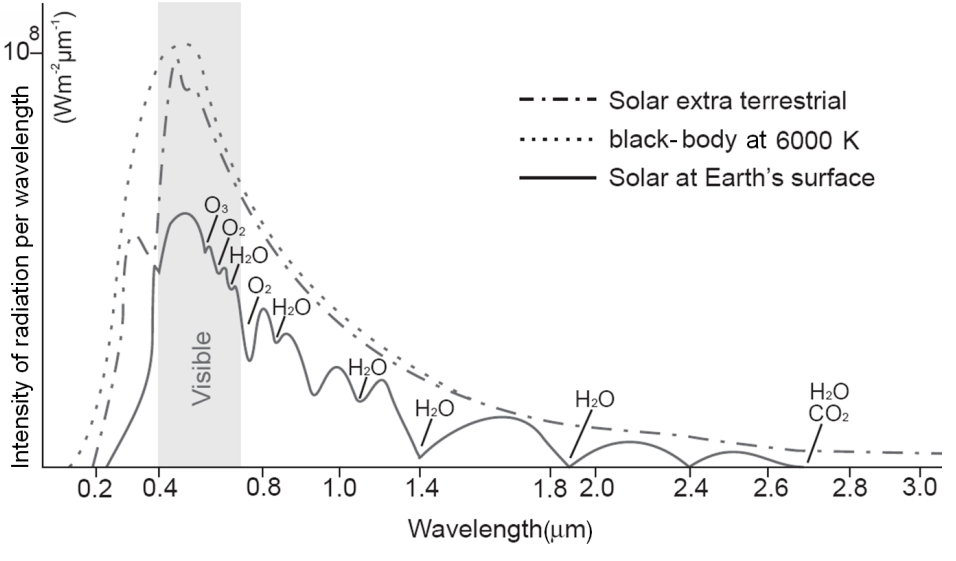


Figure 2.10: Radiation curves of the Sun and a black-body at the Sun's temperature.

2.3.2 Atmospheric scattering

Atmospheric scattering occurs when the particles or gaseous molecules present in the atmosphere cause the EM radiation to be redirected from its 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 radiant energy travels through the atmosphere. On a clear day the colours are bright, crisp. 95% of the sunlight detected by our eyes, or a comparable remote sensor is energy reflected from objects, 5% is light scattered in the atmosphere. On a cloudy or hazy day colours are faint, most of the energy hitting our eyes is scattered light. We may distinguish three types of scattering according to the size of particles in the atmosphere causing it. They are of different relevance to RS and sometimes referred to by name.

Rayleigh scattering

Rayleigh scattering dominates where electromagnetic radiation interacts with particles that are smaller than the wavelengths of light. Examples of these particles are tiny specks of dust, and nitrogen (NO_2) and oxygen (O_2) molecules. Light of shorter wavelength (blue) is scattered more than longer wavelength light (red), see Figure 2.11.

Figure 2.11: 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. At daytime the solar energy travels the shortest distance through the atmosphere; Rayleigh scattering causes a clear sky to be observed as blue. At sunrise and sunset, the sunlight travels a longer distance through the Earth's atmosphere before reaching us. All the radiation of shorter wavelengths is scattered after some distance and only the longer wavelengths reach the Earth's surface. As a result we do not see a blue but an orange or red sky (Figure 2.12).

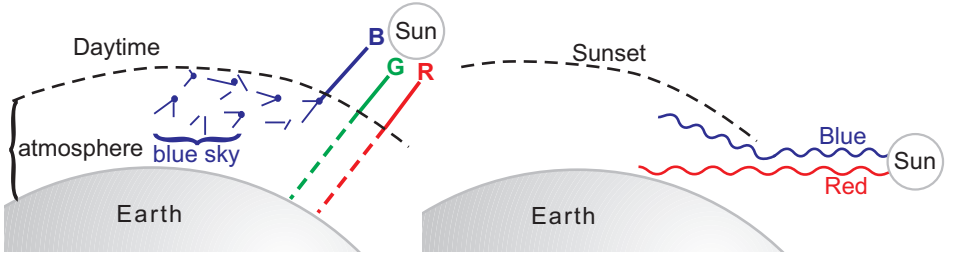


Figure 2.12: Rayleigh scattering causes us to perceive a blue sky during daytime and a red sky at sunset.

Rayleigh scattering disturbs RS in the visible spectral range from high altitudes. It causes a distortion of spectral characteristics of the reflected light when compared to measurements taken on the ground: due to Rayleigh scattering the shorter wavelengths are overestimated. In colour photos taken from high altitudes, it accounts for the blueness of these images. In general, Rayleigh scattering diminishes the "crispness" of photos and thus reduces their interpretability. Similarly, Rayleigh scattering has a negative effect on digital classification using data from multispectral sensors.

Mie scattering occurs when the wavelength of the EM radiation is similar in size to the atmospheric particles. The most important cause of Mie scattering are

Mie scattering

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 it dominates under overcast cloud conditions. Mie scattering influences the spectral range from the near-UV up to mid-IR, and has a greater effect on radiation of longer wavelengths than Rayleigh scattering.

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 the wavelength within the optical range. The most prominent example of non-selective scattering is that we see clouds as white bodies. A cloud consists of water droplets; since they scatter light of every wavelength equally, a cloud appears white. A remote sensor like our eye cannot "see through" clouds. Moreover, clouds have a further limiting effect on optical RS: clouds cast shadows (Figure 2.13).

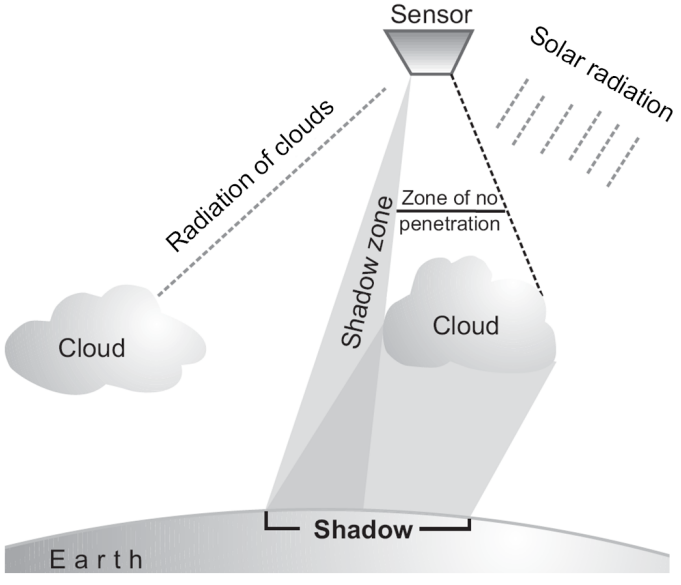


Figure 2.13: Direct and indirect effects of clouds in optical remote sensing.

2.4 Energy interactions with the Earth's surface

In "land and water applications" of remote sensing we are most interested in reflected solar energy because this tells us much about surface characteristics. Reflection occurs when radiant energy bounces off the target and is thus redirected. Not all solar energy is reflected on water and land surfaces; some of it may be absorbed and some even transmitted (through a water body to, *eg*, the sea floor). The proportion of reflected - absorbed - transmitted energy will vary with wavelength and material type. Two types of reflection, which represent the two extremes of the way in which energy is reflected by a target, are 'specular reflection' and 'diffuse reflection' (Figure 2.14). In the real world, usually a combination of both types is found.

Reflection

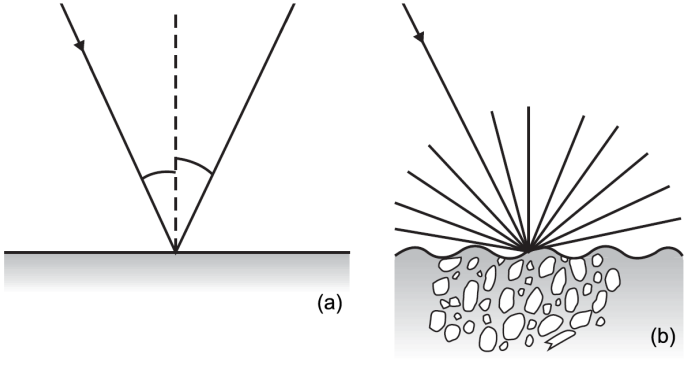


Figure 2.14: Schematic diagrams showing (a) specular and (b) diffuse reflection.

- *Specular reflection*, or mirror-like reflection, typically occurs when a surface is smooth and (almost) all of the energy is directed away from the surface in a single direction. Specular reflection can be caused, for example, by

a water 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, diffusely, or both depends on the surface roughness relative to the wavelength of the incident radiation.

2.4.1 Spectral reflectance curves

We can establish for each material type of interest a *reflectance curve*. Such a curve shows the portion of the incident energy that is reflected as a function of wavelength (expressed as percentage; see Figure 2.15). Remote sensors are sensitive to (narrow) ranges of wavelengths, not just to one particular λ, *eg*, the 'spectral band' from $\lambda = 0.4$ μm to $\lambda = 0.5$ μm. The spectral reflectance curve can be used to estimate the overall reflectance in such bands by calculating the mean of reflectance measurements in the respective ranges. Reflectance measurements can be carried out in a laboratory or in the field using a field spectrometer. Reflectance curves are typically collected for the optical part of the electromagnetic spectrum and large efforts are made to store collections of typical curves in 'spectral libraries'. In the following subsections the reflectance characteristics of some common land cover types are discussed.

Reflectance measurements

Vegetation

The reflectance characteristics of vegetation depend on the properties of the leafs, including the orientation and the structure of the leaf canopy. The amount of energy reflected for a particular wavelength depends on leaf pigmentation, leaf thickness and composition (cell structure), and on the amount of water in the leaf tissue. Figure 2.15 shows an ideal reflectance curve of healthy vegetation. In the visible portion of the spectrum, the reflection of the blue and red components of incident light is comparatively low, because these portions are absorbed by the plant (mainly by chlorophyll) for photosynthesis; the vegetation reflects relatively more green light. The reflectance in the NIR range is highest, but the amount depends on leaf development and cell structure. In the SWIR range, the reflectance is mainly determined by the free water in the leaf tissue; more free water results in less reflectance. The wavelength ranges around 1.45 and 1.95 μm are, therefore, called water absorption bands. The plant may change colour when the leafs dry out, for instance during the harvest time of the crops (*eg*, to yellow). At this stage there is no photosynthesis, which causes reflectance in the red portion of the spectrum to become higher. Also, the leafs will dry out, resulting in a higher reflectance of SWIR, whereas the reflectance of NIR may decrease. As a result, optical remote sensing can provide information about the type of plant and also about its health condition.

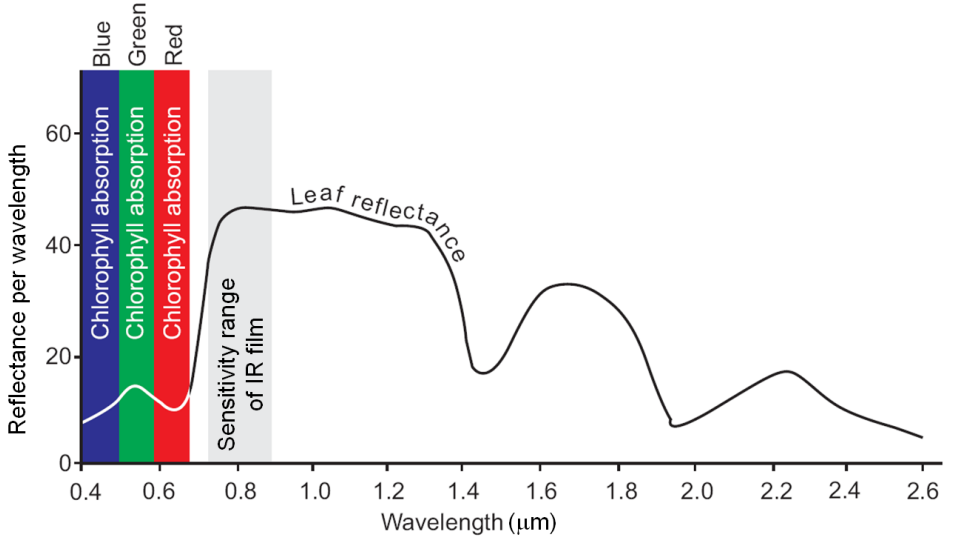


Figure 2.15: An idealized spectral reflectance curve of a healthy vegetation.

Bare soil

Reflection from bare soil depends on so many factors that it is difficult to give one typical soil reflectance curve. The main factors influencing the reflectance are soil colour, moisture content, the presence of carbonates, and iron oxide content. Figure 2.16 gives the 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 in the range 0.5 to 1.3 μm and dips at 1.45 μm and 1.95 μm. At these dips we have again water absorption bands; they are caused by the presence of soil moisture. The iron-dominated soil (e) has quite a different reflectance curve, which can be explained by the iron absorption dominating at longer wavelengths.

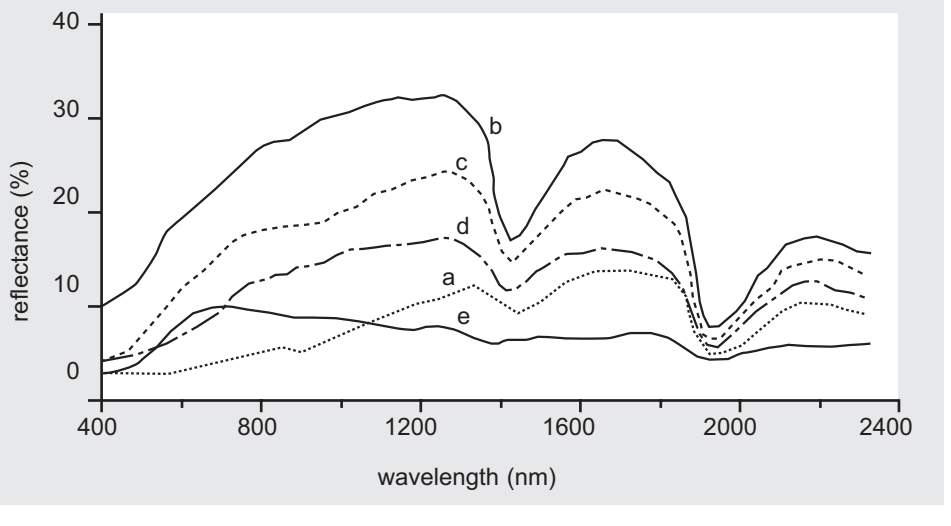


Figure 2.16: Spectral reflectance of five mineral soils, (a) organic dominated, (b) minimally altered, (c) iron altered, (d) organic affected and (e) iron dominated (from [16]).

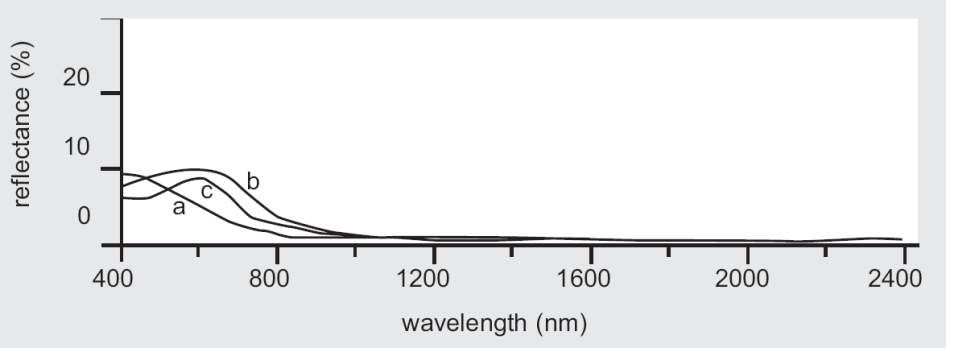


Figure 2.17: Typical effects of chlorophyll and sediments on water reflectance: (a) ocean water, (b) turbid water, (c) water with chlorophyll (from [16]).

Water

Compared to vegetation and soils, water has a lower reflectance. Vegetation may reflect up to 50%, soils up to 30–40%, while water reflects at most 10% of the incident energy. Water reflects EM energy in the visible range and a little in the NIR range. Beyond 1.2 μm all energy is absorbed. Spectral reflection curves of different types of water are given in Figure 2.17. Turbid (silt loaded) water achieves the highest reflectance. Water containing plants has a pronounced reflectance peak for green light because of the chlorophyll of the plants.

2.5 Sensing of EM energy

The review of properties of EM radiation has shown that different forms of radiant energy can provide us with different information about terrain surface features and that different applications of earth observation are likely to benefit from sensing in different sections of the useful EM radiation range. A geoinformatics engineer who wants to discriminate objects for topographic mapping prefers to use an optical sensor operating in the visible range. An environmentalist who needs to monitor heat losses of a nuclear power plant will use a sensor that detects thermal emission. A geologist interested surface roughness because it enlightens him on rock types will rely on microwave sensing. Different demands combined with different technical solutions have resulted in a manifold of sensors. In this section we will classify the various remote sensors and treat their commonalities. Peculiarities will then be treated in the respective later chapters.

2.5.1 Sensing properties

A *remote sensor* is a device that detects EM energy, quantifies it, and usually records it, in an analogue or digital way. It may also transmit recorded data (to a receiving station on the ground). Many sensors used in earth observation detect reflected solar energy. Others detect the energy emitted by the Earth itself. However, there are some bottlenecks. The Sun is not always shining brightly and there are regions on the globe almost permanently under cloud cover. There are regions with seasons of very low sun elevation, so that objects cast long shadows over long periods. There is night time with only "atmospheric night-glow" and perhaps moonlight. Sensors detecting reflected solar energy are useless at night and face problems under unfavourable season and weather conditions. Sensors detecting emitted terrestrial energy do not directly depend on the Sun as a source of illumination; they can be operated any time. The Earth's emission, we have learned, is only at longer wavelengths because of the relatively low surface temperature and long EM waves are not very energetic, thus more difficult to sense.

Bottlenecks of sensing

Luckily we do not have to rely on solar and terrestrial radiation only. We can build instruments, which emit EM energy and then detect the energy returning from the target object or surface. Such instruments are called *active sensors* as opposed to passive ones, which measure solar or terrestrial energy (Figure 2.18). An example of an active sensor is a laser rangefinder, a device we can buy for a few Euros in any DIY-shop. Another very common active sensor is a camera with a flash light (operated at night). A camera without flash is a passive sensor. The main advantages of active sensors are that they can be operated day and

Active versus passive remote sensing

night, are less weather dependent, and have a controlled illuminating signal, which is less affected by the atmosphere. Laser and radar instruments are the most prominent active sensors for GDA.

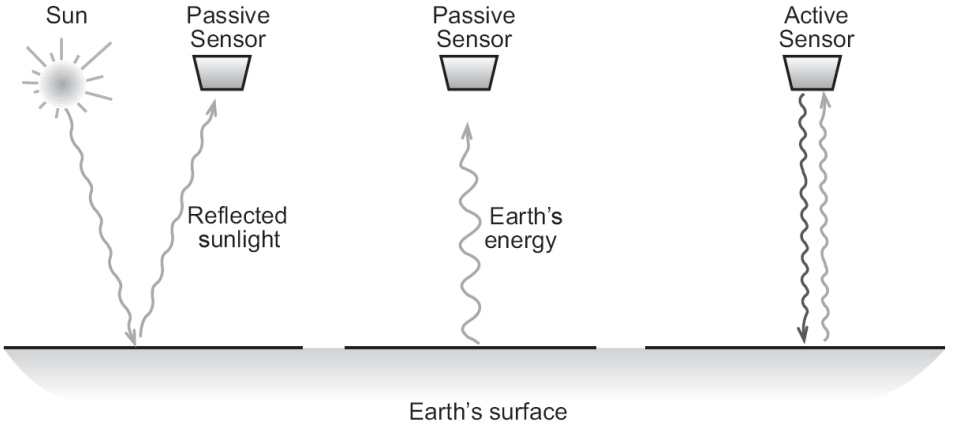


Figure 2.18: A remote sensor measures reflected or emitted energy. An active sensor has its own source of energy.

Most remote sensors measure either an intensity change or a phase change of EM radiation. Some - like a simple laser rangefinder - only measure the elapse time between sending out an energy signal and receiving it back. Radar sensors may measure both intensity and phase. Phase measuring sensors are used for precise ranging (distance measurement), *eg*, by GPS 'phase receivers' or continuous-wave laser scanners. The intensity of radiation can be measured via the photon energy striking the senor's sensitive surface.

Intensity or phase

By considering the following, you can relate the intensity measure of reflected energy to Figure 2.8 and link the Figures 2.15 to 2.17. When sensing reflected

Measuring radiance

light the radiance at the detector is the radiance at the Earth's surface attenuated by atmospheric absorption plus the radiance of scattered light:

$$L = \frac{\rho E \tau}{\pi} + skyradiance \tag{2.4}$$

where L is the total radiance at the detector, E is the irradiance at the Earth's surface, ρ is the terrain reflectance, and τ is the atmospheric transmittance. The radiance at the Earth's surface depends on the irradiance (the intensity of the incident solar radiation) and the terrain surface reflectance. The irradiance in turn stems from direct sunlight and diffuse light, the latter caused by atmospheric scattering and the more so on a hazy day (see Figure 2.19). This indicates why you should study radiometric correction (Section 5.3 and Section 11.4) to better infer on surface features.

Spectral band

The radiance is observed for a 'spectral band', not for a single wavelength. A *spectral band* or wavelength band is an interval of the EM spectrum for which the average radiance is measured. Sensors like a panchromatic camera, a radar sensor, or a laser scanner only measure in one specific band while a multispectral scanner or a digital camera measures in several spectral bands at the same time. Multispectral sensors have several 'channels', one for each spectral band. Figure 2.20 shows spectral reflectance curves together with the spectral bands of some popular satellite based sensors. Sensing in several spectral bands simultaneously allows us to relate properties that show up well in specific spectral bands. For example, reflection characteristics in the spectral band 2 to 2.4 μm (as recorded by Landsat-5 TM channel 7) tell us something about the mineral composition of the soil. The combined reflection characteristics in the red and

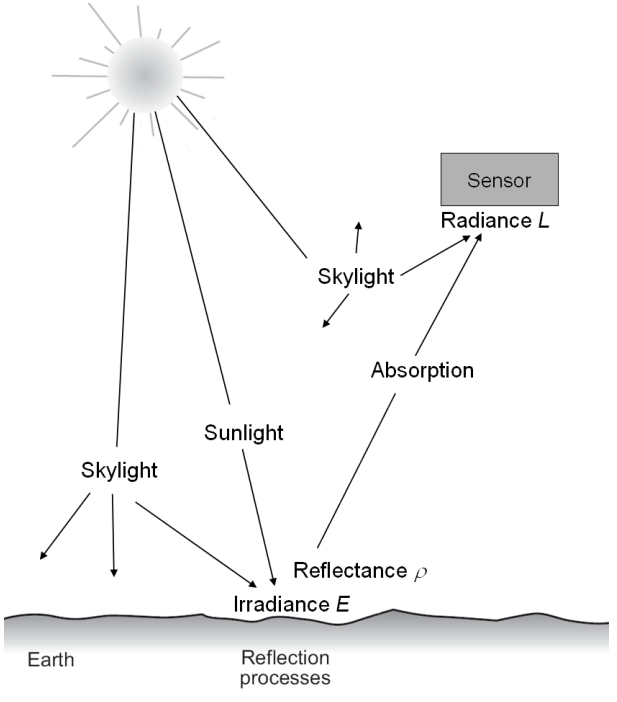


Figure 2.19: Radiance at the sensor, adapted from [15].

NIR bands (from Landsat-5 TM channels 3 and 4) can tell us something about biomass and plant health.

Landsat MSS (MultiSpectral Scanner), the first civil spaceborne earth observation sensor had detectors for three rather broad spectral bands in the visible part of the spectrum, with a width of 100 nm. A hyperspectral scanner uses detectors for many more but narrower bands, which may be as narrow as 20 nm or even

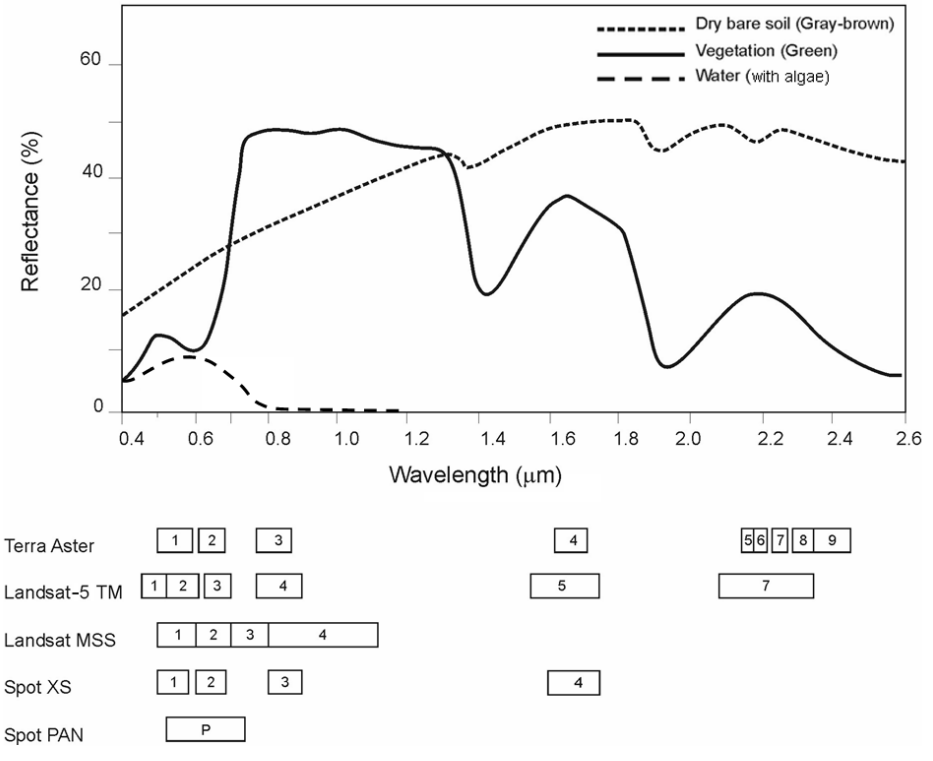


Figure 2.20: Spectral reflectance curves and spectral bands of some multispectral sensors.

less. We say a hyperspectral sensor has a higher 'spectral resolution' than a multispectral one. A laser instrument can emit (and detect) almost monochrome radiation, with a wavelength band not wider than 10 nm. A camera loaded with panchromatic film or a spaceborne electronic sensor with a panchromatic channel (such as SPOT PAN or WorldView-1) records the intensity of radiation of a broad spectral band, covering the entire visible range of the EM spectrum.

Spectral resolution

Panchromatic - which stands for "across all colours" - recording is comparable with the function of the 120 million rods of a human eye. They are brightness detectors and cannot sense colour.

Photographic detector

In a camera loaded with panchromatic film ('black-and-white film'), the detection of radiant energy is done by the silver halide crystals of the light-sensitive emulsion. Silver halide grains turn to silver metal when exposed to light, the more so the higher the intensity of the incident light. Each light ray from an object/scene triggers a chemical reaction of some grain. This way energy variations within a scene are detected and an image of the scene is created at the time of exposure. The obtained record is only a latent image; it takes film development to turn it into a photograph.

AD conversion

Digital cameras or multispectral scanners are examples of sensors, which use electronic detectors instead of photographic ones. An electronic detector (CCD, CMOS, photodiode, solid state detector, etc) is made of semiconductor material. The detector accumulates charge by converting the photons incident upon its surface to electrons. It was Einstein, who won the Nobel price for discovering and explaining that there is an emission of electrons when a negatively charged plate of light-sensitive (semiconductor) material is subject to a stream of photons. The electrons can then be made to flow as a current from the plate. So the charge can be converted to a voltage (electrical signal). The charge collected is proportional to the radiance at the detector (the amount of energy deposited in the detector). The electrical signal is 'sampled' and quantized (Figure 2.21). This process is called A/D conversion; the output is a digital number (DN), which is recorded. The DN is an integer number within a fixed range. Older remote sen-

sors have used 8 bits recording, which allows a differentiation of radiance into $2^8 = 256$ levels (*ie*, DNs in the range 0 to 255). The recent WorldView-1 sensor (launched in 2007) records with a *radiometric resolution* of 11 bits $(2^{11} = 2048)$. ASTER records for the visible spectral band with 8 bits, for the thermal infrared band with 12 bits. A higher radiometric resolution requires more storage capacity but has the advantage of offering data with higher information content (see Figure 2.22).

Radiometric resolution

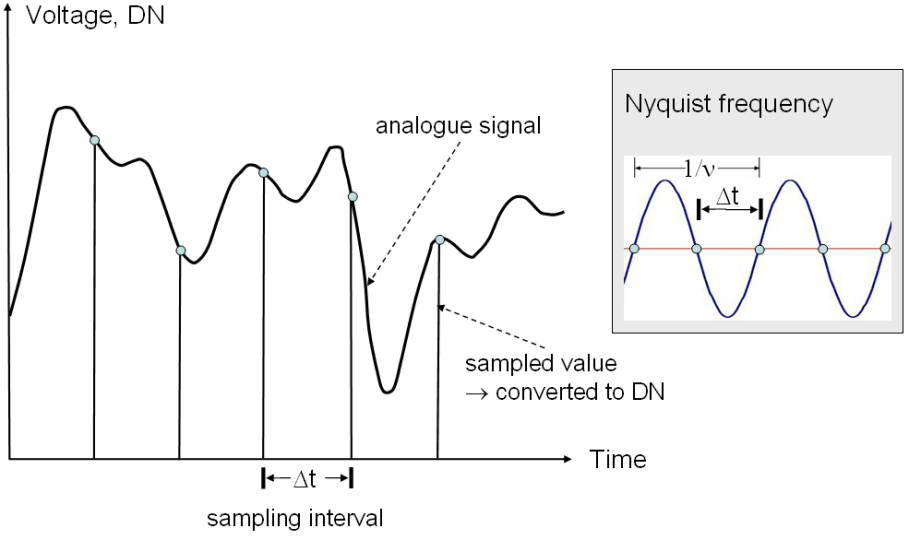


Figure 2.21: Illustration of sampling as understood in signal processing. The sampling interval should be smaller than half the period of the highest oscillation contained in the signal (which is a challenge to the chip producer).

A digital panchromatic camera has an array of detectors instead of silver halide crystals suspended in gelatine on a polyester base in a film camera. Each detector (*eg*, a CCD, which stands for charge-coupled device) is of very small size,

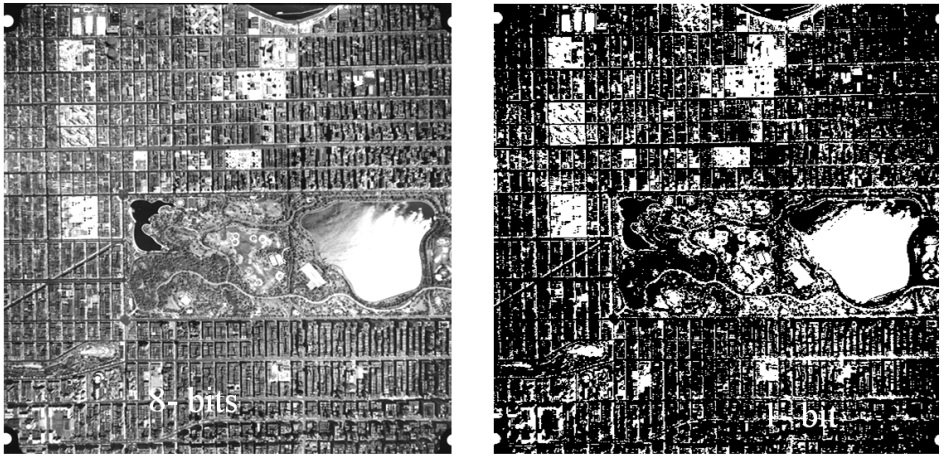


Figure 2.22: 8 bits versus 1 bit radiometric resolution.

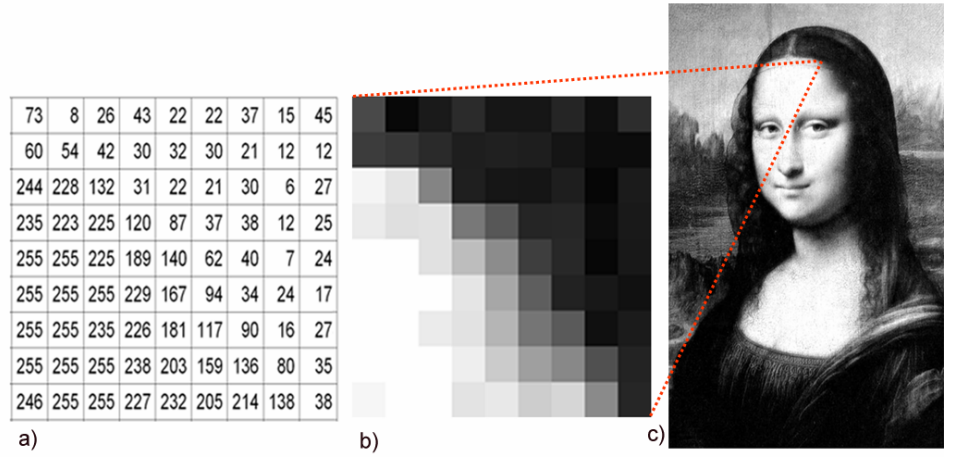


Figure 2.23: Digital image - (a) data from sensors, (b) display by grey values, (c) image zoomed out.

in the order of 9 µm by 9 µm. Spaceborne cameras use larger detectors than aerial cameras to collect enough photons despite of the long distance to the Earth. At the moment of exposure each detector yields one DN, so in total we obtain a data set, which represents an image similar to the one which is created by exciting photographic material in a film camera. When arranging the DNs in a two-dimensional array we can readily visualize them as grey values (Figure 2.23). We refer to the obtained "image" as *digital image* and to a sensor producing digital images as *imaging sensor*. The array of DNs represents an image in terms of discrete picture elements, called *pixels*. The value of a pixel - the DN - corresponds to the radiance of the light reflected from the small ground area viewed by the respective detector (according to Formula 2.4). The smaller the detector the smaller will be the area on the ground which corresponds to one pixel. The size of the 'ground resolution cell' is often referred to as 'pixel size on the ground'. Early digital consumer cameras had 2×10^6 CCDs per spectral band (named 2 megapixels cameras); today we can get for the same price a 10 megapixels camera. The latter has much smaller CCDs so that they fit on the same board, with the consequence that an image can reveal much more detail; we say the 'spatial resolution' of the image is higher.

Imaging, spatial resolution

A digital consumer camera does not record intensity values for a single (panchromatic) spectral band, but for three bands simultaneously, namely for red, green, and blue light in order to obtain colour images. This is comparable to our eyes: we have three types of cones, one for each primary colour. The data set obtained for one shot taken with the camera (the 'image file'), therefore, contains three separate digital images (Figure 2.24). Multispectral sensors record in as many as 14 bands simultaneously (*eg*, ASTER). For convenience, a single digital image is then often referred to as 'band' and the total image file as 'multi-band image'.

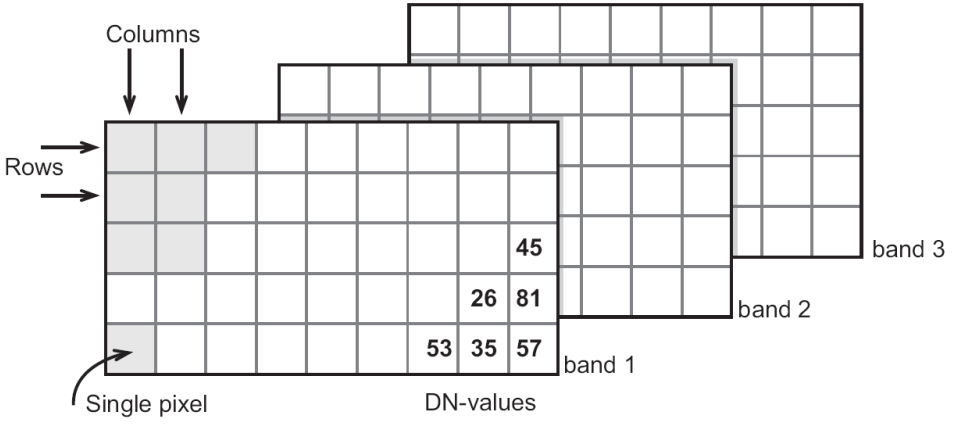


Figure 2.24: An image file comprises a digital image for each of the spectral bands of the sensor. For every band the DN-values are stored in a row-column arrangement.

Storage media

Various storage media are used for recording the huge amount of data produced by electronic detectors: solid state media (such as memory cards as used for consumer cameras), magnetic media (disk or tape), and optical discs (some video cameras). Satellites usually have a couple of recorders on-board; light sensor systems often transmit data to ground receiving stations at night. However, data can also be transmitted directly to a receiving station using satellite communication technology. Airborne sensors often use the hard-disk of a laptop computer as recording device.

2.5.2 Classification of sensors

We can classify and label remote sensors in different ways. According to what our prime interest in earth observations may be - geometric properties, spectral differences, or intensity distribution of an object or scene - we can distinguish three salient types: altimeters, spectrometers, and radiometers. Laser and radar altimeters are non-imaging sensors, providing us with information on elevation of water and land surfaces.

Altimeter

Thermal sensors such as the channels 3 to 5 of NOAA's AVHRR or the channels 10 to 14 of Terra's ASTER are called (imaging) radiometers. Radiometers measure radiance and typically sense in one broad spectral band or in only a few bands, but with high radiometric resolution. Panchromatic cameras and passive microwave sensors are other examples. The spatial resolution depends on the wavelength band of the sensor. Panchromatic radiometers can have a very high spatial resolution, whereas microwave radiometers have a low spatial resolution because of the low energy level in this spectral range. The scatterometer referred to in Chapter 1 is a non-imaging radiometer. Radiometers have a wide application range; they are used to detect forest/bush/coal fires, determine soil moisture and plant response, monitor ecosystem dynamics, analyse energy balance across land and sea surfaces, etc.

Radiometer

Spectrometers measure radiance in many (up to 224) narrow, contiguous spectral bands, thus have a high spectral resolution. Their spatial resolution is moderate to low. The prime use of imaging spectrometers is identifying surface materials - from mineral composition of soils to suspended matter concentration

Spectometers

of surface water and chlorophyll contents. There are also androgynies: spectro-radiometers, imaging laser scanners, Fourier spectrometers, etc.

We can also group the manifold of remote sensors used for GDA with respect to the spectral domains in which they operate (Figure 2.25). In the following you will find a short description of the groups and the reference to the chapter where they are treated in more detail.

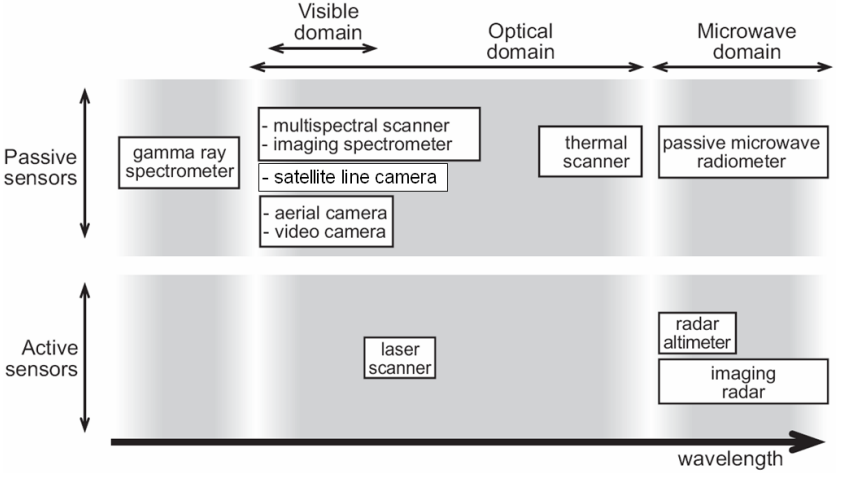


Figure 2.25: Overview of the sensors that are exposed in this book.

- Gamma ray spectrometers have their main application in mineral explo- Gamma ray sensors

ration. Gamma ray surveys are treated in Chapter 14.

- Aerial film cameras have been the workhorse for many decennia. They find their prime use today in large scale topographic mapping, cadastral mapping, orthophoto production for urban planning, etc. They are discussed in Chapter 9. Film cameras

- Digital aerial cameras are not as quickly conquering the market as digital cameras became popular on the consumer market. They use CCD arrays instead of film and are treated together with optical scanners in Chapter 4. Line cameras operated from satellites have very similar properties. Digital cameras

- Digital video cameras are not only used to record everyday motion pictures. They are also used in aerial earth observation to provide low cost (and low resolution) images for mainly qualitative purposes, for instance to provide visual information of the area covered by "blind" airborne laser scanner data. Handling images stemming from video cameras is similar to dealing with images from digital 'still' cameras and is not further explicitly discussed in this book. Video cameras

- Multispectral scanners are mostly operated from satellites and other space vehicles. The essential difference with satellite line cameras is in the imaging/optical system: they use a moving mirror to 'scan' a line (*ie*, a narrow strip on the ground) and a single detector instead of recording intensity values of an entire line at one instant by an array of detectors. They are treated in Chapter 4. Figure 2.26 shows an image obtained by combining the images of the Landsat TM channels 4, 5, and 7, which are displayed in Multispectral scanners

red, green and blue respectively. Chapter 5 explains how to produce such a 'false colour image'.

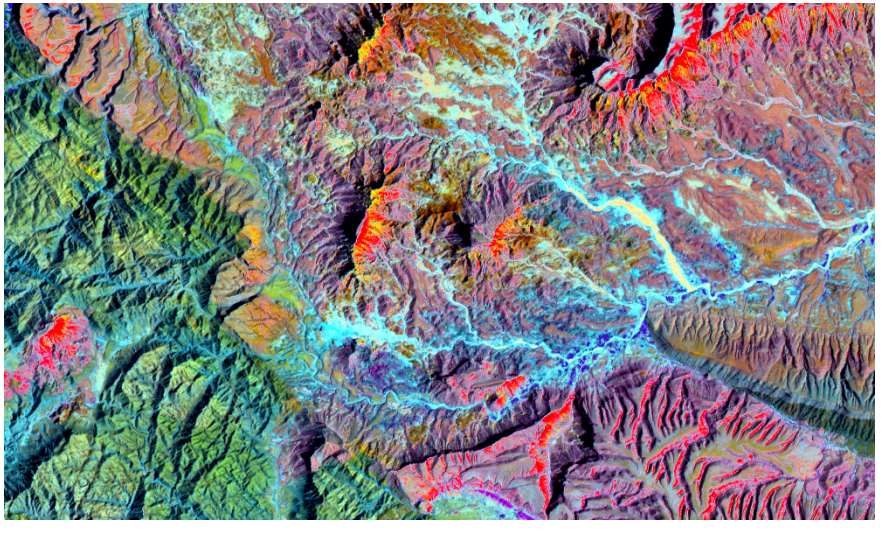


Figure 2.26: Landsat-5 TM false colour composite of an area of 30 km by 17 km.

Imaging spectrometers

- Hyperspectral scanners are imaging spectrometers with a scanning mirror. They are touched upon in Chapter 4 and treated in detail in Chapter 13.

Thermal scanners

- Thermal scanners are placed here in the optical domain for the sake of convenience. They exist as special instruments and as a component of multispectral radiometers. They are included in Chapter 4 and the processing of thermal data is treated in Chapter 12. Thermal scanners provide us with data, which can directly be related to object temperature. Figure 2.27 shows an example of a thermal image acquired by an airborne thermal scanner flown at night.

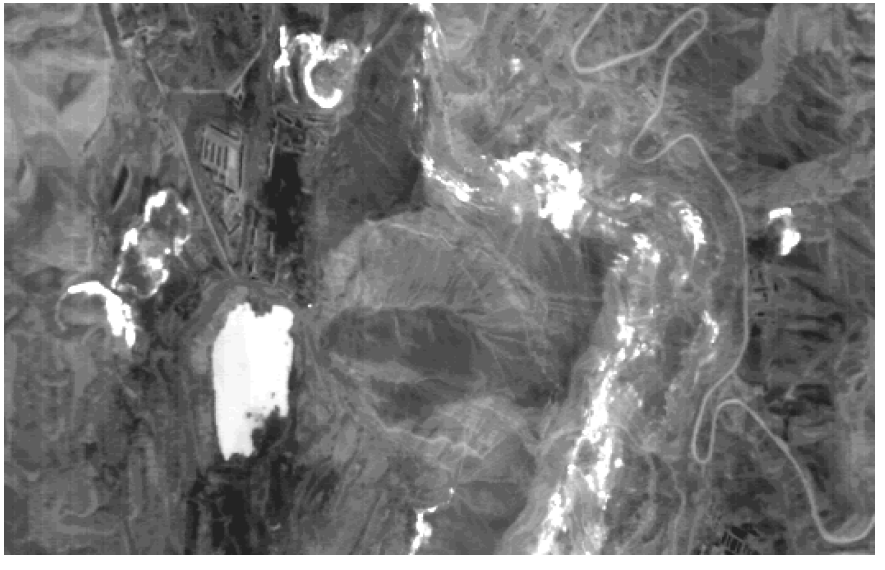


Figure 2.27: 'Thermal image' of a coal mining area affected by underground coal fires. Darker tones represent colder surfaces, while lighter tones represent warmer areas. Most of the warm spots are due to coal fires, except for the large white patch, which is a lake. Apparently, at that time of the night, the temperature of the water is higher than the temperature of the land. The scene is approximately 4 km across.

- Passive microwave radiometers detect emitted radiation of the Earth's surface in the 10 to 1000 mm wavelength range. Main applications are in mineral exploration, monitoring soil moisture changes, and snow and ice detection. Microwave radiometers are not further discussed in this book.

Microwave radiometers

- Laser scanners are the scanning variant of laser rangefinders and laser altimeters (as the one on ICESat). Laser scanners measure the distance from the laser instrument to many points of the target in "no time" (*eg*, 150,000 in one second). Laser ranging is often referred to as LIDAR (LIght Detection And Ranging). The prime application of airborne laser scanning is to create high resolution digital surface models and digital terrain relief mod-

Laser scanners

els (see Chapter 6). We can create a digital terrain relief model (DTM) also from photographs and similar panchromatic images; however, because the properties of laser radiation, airborne laser scanning has important advantages in areas of dense vegetation, sand deserts and coastal areas. Surface modelling is of interest to many applications, for example, biomass estimation of forests, volume calculation in open pit mining (see Figure 2.28), flood plain mapping, 3D modelling of cities, etc. Laser scanning is elaborated in Chapter 10.

Figure 2.28: Pictorial representation of a digital surface model of the open pit mine on the Sint Pietersberg, the Netherlands. The size of the pit is roughly 2 km by 1 km. Clearly visible is the terraced rim of the pit. The black strip near the bottom of the image is the river Maas. Courtesy Survey Department, Rijkswterstaat.

- Imaging radar (RAdio Detection And Ranging) operates in the spectral domain from 10 to 1000 mm. Radar instruments are active sensors and because of the used wavelength range they can provide us with data day and night, and under all weather conditions. Radar waves can penetrate

clouds, only heavy rainfall affects imaging somewhat. One of the applications is, therefore, mapping of areas that are under permanent cloud cover. Figure 2.29 shows an example of a SAR (Synthetic Aperture Radar) image from the ERS-1 satellite. Radar data from air or space can also be used to create surface models. Radar imaging has a peculiar geometry and processing raw radar data is not simple. Radar is treated in Chapter 10.

Imaging radar

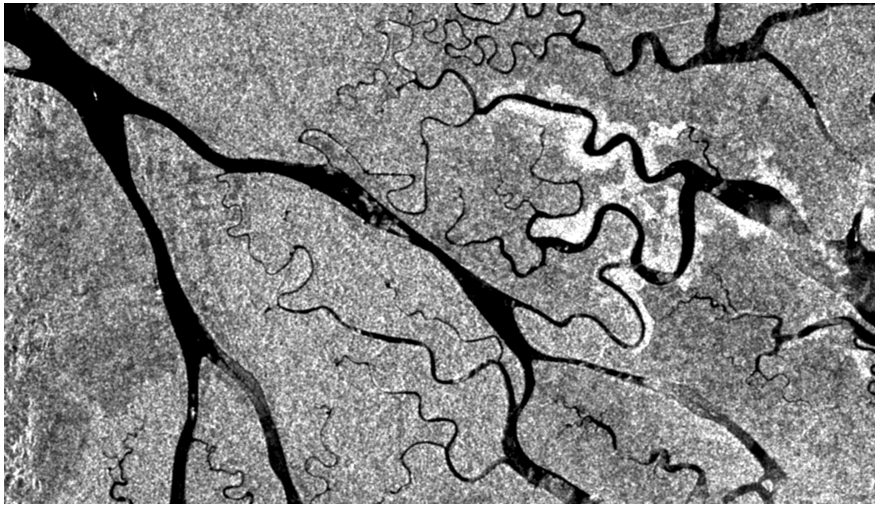


Figure 2.29: ERS-1 SAR image of the Mahakam Delta, Kalimantan. The image shows different types of land cover. The river itself is black. The darker patch of land the the left is the inland tropical rainforest. The rest os a mixed forest of Nipa Palm and Mangrove on the delta. The right half of the image shows light patches, where the forest is partly cleared. The image covers and area of 30 km x 15 km.

- Radar altimeters are used to measure elevation profiles of the Earth's surface parallel to the satellite orbit (see Section 4.2). They operate in the 10 to 60 mm range and allow us to calculate elevation with an accuracy of 20 to 50 mm. Radar altimeters are useful for measuring relatively smooth sur-

faces. Currently there are four spaceborne radar altimeters in operation; Envisat (see Section 4.6) carries one of them. Figures 1.4 and 14.2 show examples of radar altimeter products.

Radar altimeters

- For the sake of completeness we include here a further active sensor: sonar. Sonar stands for SOund NAvigation Ranging. It is used for mapping river beds, sea floors, detect obstacles underwater, etc. It works by emitting a small burst of sound from a ship. The sound is reflected off the bottom of the body of water. The time that it takes for the reflected pulse to be received corresponds to the depth of the water. More advanced systems also record the intensity of the return signal, thus giving information about the material on the sea floor. In its simplest form, the sonar looks straight down, and is operated very much like a radar altimeter. The body of water will be traversed in paths like a grid; not every point below the water surface will be monitored. The distance between data points depends on the ship's speed, the frequency of the measurements, and the distance between the adjacent paths. One of the most accurate systems for imaging large areas of the ocean floor is called the side scan sonar. It is an imaging system, which works somewhat similar to side looking airborne radar (see Chapter 10). The images produced by side scan sonar systems are highly accurate and can be used to delineate even very small (< 1 cm) objects. Using sonar data we can produce contour maps of sea floors and other water bodies, which find their application in ship navigation and water discharge studies.

Sonar

Summary

Remote sensing is based on detecting electromagnetic (EM) energy. EM energy travels through space in the form of sine waves with interacting electrical and magnetic oscillation. EM radiation can be modelled either by waves or by a stream of 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 micrometres (μm, 10^{-6} m) or nanometres (nm, 10^{-9} m). The frequency is the number of cycles of a wave passing a fixed point in space in a second; it 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 absolute zero (0 K) emits EM energy due to molecular agitation. Matter that is capable of absorbing and re-emitting all EM energy received is known as a black-body. All matter with a certain temperature emits radiant energy 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 sensor is a function of the interactions at the Earth's surface and the energy interactions in the atmosphere.

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 absorbed, scattered, transmitted, and reflected. The most efficient absorbers of solar energy in the at-

mosphere are ozone molecules, water vapour, and carbon dioxide. 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. All types of scattering are disturbing to RS of land and water surfaces.

When solar energy reaches the Earth's surface, three fundamental energy interactions are possible: absorption, transmission, and reflectance. Specular reflection occurs when a surface is smooth and the incident 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. The spectral property of a material's reflection is presented by a reflectance curve.

In earth observation, sensed radiation of wavelengths up to 3 μm is predominately reflected solar energy while infrared radiation above 3 μm can be mainly attributed to emitted energy, namely terrestrial heat. Because reflected EM energy follows the laws of optics (reflection, refraction, transmitting and focusing of rays by lenses or mirrors, etc) sensors operating in this range are often referred to as optical remote sensors.

We use active and passive sensors for GDA. Passive sensors depend on an external source of energy; in particular the Sun. Active sensors have their own source of energy. Most of the remote sensors discussed in this book are imaging sensors. They record intensity levels, which correspond to radiances of reflected or emitted EM energy of the target area. An electronic sensor "measures intensity"

by detecting photons, converting those to electrons and the collected charge to an electrical signal. The analogue electrical signal is sampled and converted to a digital number (DN). The increment of quantizing determines the radiometric resolution. Most common today is storing a DN in 8 bits, which implies a possible DN range from 0 to 255. Remote sensors measure in discrete spectral bands; those measuring in many narrow bands are said to have high spectral resolution. An imaging sensor covers a target area by what we call a RS image. A RS image can be a single-band image or a multi-band image. A digital image is a single-band image composed of discrete picture elements (pixels). The smaller the pixels size the higher the spatial resolution of the image. Remote sensors make use of various storage media. Classifying the manifold of remote sensors helps us to find the right sensor for a particular application. When classifying with respect to function, we distinguish as main groups altimeters, radiometers, and spectrometers. When classifying with respect to the spectral domain we can distinguish two super-classes: optical sensors (including TIR sensors) and microwave sensors.

Additional reading

[15]

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 disturbances?
2. How important are laboratory spectra for understanding remote sensing images?

These are typical exam questions:

1. List and describe the two models used to explain electromagnetic energy.
2. How are wavelength and frequency related to each other (give a formula)?
3. What is the electromagnetic spectrum?
4. Describe three types of atmospheric scattering.
5. What are the specific energy interactions that take place when EM energy from the Sun hits the Earth's surface?
6. Give a definition of 'atmospheric window' in your own words.
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 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

Spatial referencing

3.1 Introduction

How can we relate data acquired by a sensor in the air or in space to a position on the ground? Chapter 6 will explain it. It takes spatial reference frames that we can position objects on the Earth's surface in a terrestrial 3D coordinate system and subsequently derive 2D map coordinates as needed for showing geospatial data on computer screens or paper maps. Establishing spatial reference frames is the subject of geodesy. Geodesy is an exact science, different from geography, which is descriptive. Geodetic positioning requires both sound knowledge of physics and mathematics. You have to learn the basics of determining position on the Earth's surface in order to understand 'georeferencing' of remote sensing images and 'direct sensor orientation'. Moreover, you should be able to assess how accurately you have geometrically corrected a remote sensing image. We can measure positional accuracy in various ways. Within the scope of this book it is sufficient to judge achieved accuracy in relating images to each other or to maps by the RMSE measure.

For ITC historical reasons the chapter on determining position is in the book *Principles of GIS* (PoGIS). To avoid abundant duplication, the following section only reviews the most essential concepts of spatial referencing as needed for the remainder of this book.

3.2 Review of concepts needed for RS

We know for more than 2000 years that the Earth is spherical. However, the oldest existing terrestrial globe was not made before 1474 (in Germany). RS images and (most) maps are both 2D representations of the round word but they are generated by following different projection principles. In order to obtain a map on a computer screen or a paper map, which we can fold up and put it in our pocket, we could think of cutting up a globe in the way we often peel an orange. Each resulting skin segment can more easily be flattened than trying to make the surface flat as a whole. In map projection, such a segment is called a 'zone'. Mercator's mathematical approach to projecting a spherical surface onto a flat surface was used for the maps used by Christopher Columbus. The Mercator projection is one of the oldest map projections; many more have been developed since the 16th century, each one having its specific advantages. The most widely used map projection today, however, is the Universal Transverse Mercator (UTM) projection. UTM is a 'cylindrical map projection', which can be thought of as projecting a spheroid in zones onto a cylinder and then unwinding the cylinder; see Figure 3.1.

The input to a map projection is a sphere or an ellipsoid (also referred to as spheroid). An ellipsoid is a simple mathematical shape, which nicely approximates the shape of the Earth at mean sea level and thus can serve as a horizontal reference for position on the Earth's surface. We can define parallels (to the equator) and meridians on the ellipsoid and thus the geographic coordinates latitude and longitude (in short Lat-Lon or ϕ and λ , see Figure 3.1). In the course of the years different reference ellipsoids have been defined for different parts of the World in order to obtain a locally best fitting surface. The ellipsoid for the

national coordinate system of the Netherlands and Germany is referred to as the Bessel ellipsoid. The ellipsoid for the international spatial reference system WGS84 is the GRS80 ellipsoid; it has slight different dimensions from the Bessel ellipsoid. In order to attain a best fit of a chosen ellipsoid to the local mean sea level of a region, geodesists determined in the past an optimal positioning and orientation of the ellipsoid, this way defining a local/national datum. The consequence for the GDA practice is that we find ourselves often confronted with different reference coordinate systems and that we have to apply a datum transformation. A typical example is that we obtain data by satellite positioning, *eg*, by GPS, and that we have to integrate them with data available in a national coordinate system. Geographic coordinates of a point, *eg*, at the entrance to the ITC, in the WGS84 system are different form the geographic coordinates given in the Dutch geodetic system. To apply a datum transformation, we must know the ellipsoid parameters and the datum parameters of the two systems unless we know directly the transformation parameters (as published by the national mapping agency).

RS images from satellites are provided either as 'raw images' (*ie*, without a geometric reference to positions on the ground) or in a geometrically processed form. If the latter, the coordinates of points in the image are directly map coordinates, usually UTM coordinates based on the WGS84 system.

Additional reading

You are advised to critically read the PoGIS sections on satellite-based position-

ing, which outline basic principles, characteristics of current implementations and related accuracy levels, and the section on positional accuracy - before continuing with Chapter 4.

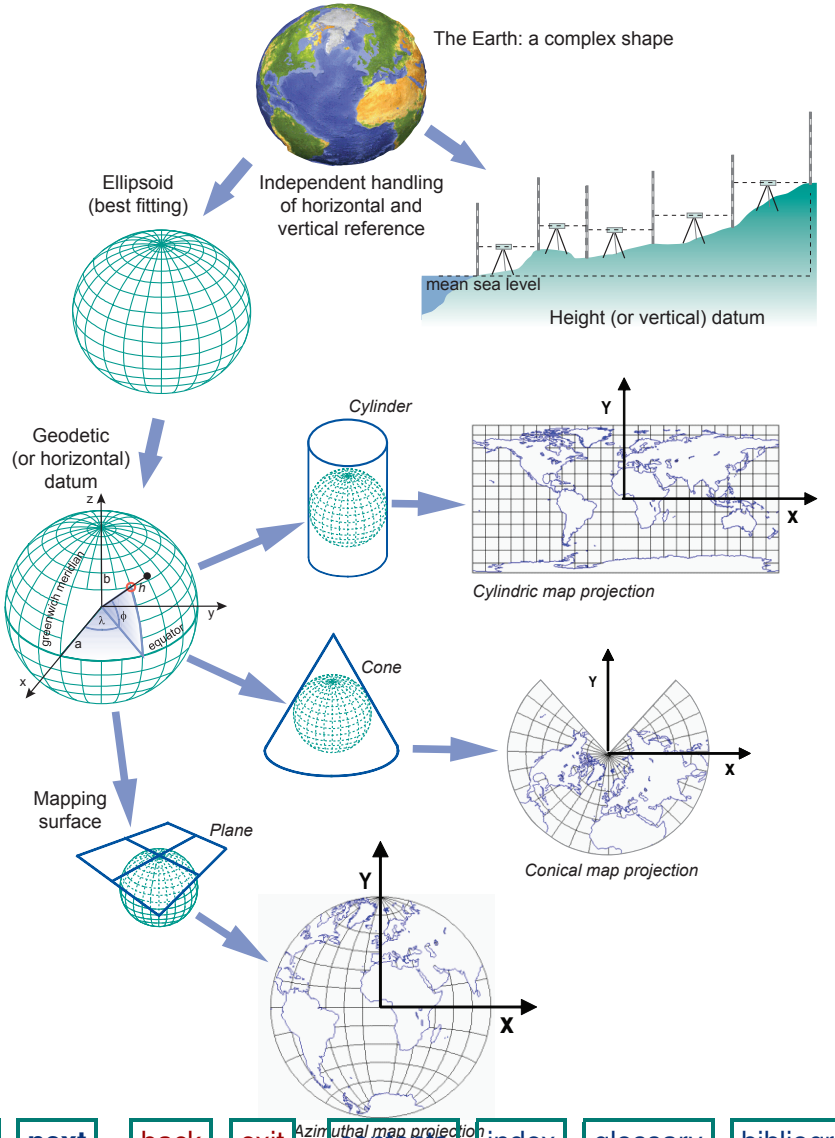


Figure 3.1: Illustration of geospatial reference systems.

Chapter 4

Platforms and passive electro-optical sensors

4.1 Introduction

In Chapter 2, the physics of sensing have been explained. This chapter discusses sensor systems and tries to discover the logic of current electro-optical sensing techno*logy*. We will first consider the characteristics of platforms used for GDA from air and space. Various aircrafts, space shuttles, space stations, and satellites are used to carry one or more sensors for earth observation. Section 4.3 will elaborate on frame cameras and line cameras, the latter being operated from air or space are also known as 'pushbroom sensors'. Optical scanners (also referred to in literature as across-track scanners or 'wiskbroom scanners') are treated in Section 4.4, covering multispectral, hyperspectral and thermal scanners. Some camera systems can provide us with 'stereo' images; therefore, Section 4.5 will give a short introduction to stereoscopy. Data from spaceborne sensors are widely available through data providers; an overview of the most popular satellite based sensors is given in Section 4.6. The sensor-platform combination determines the characteristics of the resulting data. Based on your information needs and on time and budgetary criteria, you can determine which data source is most appropriate; Section 4.7 summarizes data selection criteria.

4.2 Platforms and missions

Sensors used in earth observation can be operated at altitudes ranging from just a few centimetres above ground - using field equipment - to far beyond the atmosphere. Very often the sensor is mounted on a moving vehicle - which we call the *platform* - such as an aircraft or a satellite. Occasionally, static platforms are used. For example, we can mount a spectrometer on a pole to measure the changing reflectance of a specific crop during the day or a season.

4.2.1 Moving platforms

To gain a wider view, we use aircrafts, up to altitudes of around 20 km. Depending on the type of aerial survey, equipment weight, and survey costs we can choose from a variety of vehicles. We use fixed wing aircrafts for thermal scanning and a systematic photo-coverage for topographic mapping, land titling projects, etc. Aerial survey cameras are heavy and they are fixed to a stabilized mount in the hole in the floor of the airplane (see Section 4.3). Most survey airplanes fly lower than 8 km but higher than 500 m. They can fly as slow as 150 km/h, but even at such a speed image quality is already affected by motion blur unless the camera has a compensation device (see Chapter 9). Aerial survey cameras are highly sophisticated and expensive. Also airborne laser scanner systems used to be heavy, but now the total equipment set can already be as light as 30 kg. Laser scanners are either mounted on fixed wing aircrafts or helicopters. Helicopters can fly at low altitudes and very slowly, thus allowing for very detailed data acquisition (at high costs per square km). Small-format cameras are cheaper and lighter than large format aerial cameras. They can be mounted on micro-light airplanes for urban reconnaissance or even on kites, *eg*, for surveying and industrial area. Unmanned aerial vehicles (UAVs) are gaining popularity for observing possibly dangerous areas (*eg*, for monitoring cannabis growing in Holland or elsewhere) or to reduce costs. A special type of UVA, the High Altitude Long Endurance (HALE) vehicles bridge the gap between manned survey aircrafts and spacecrafts or satellites. These are remotely operated aircrafts of ultra light weight and load, flying for months at altitudes around 20 km. A key advantage of aerial surveys is that they can be "targeted". The survey can be undertaken at exactly the wanted time. It can also be done with exactly the wanted spatial resolution by letting the aircraft fly at the needed altitude. More-

Aircrafts

over, compared with civil satellite RS, we can acquire images of much higher spatial resolution, thus recognize objects of much smaller size. We can achieve a pixel size on the ground as small as 5 cm with current aerial cameras.

Satellites

Satellites are launched into space with rockets where they then circle round the Earth for 5 to 12 years in a predefined orbit. The choice of the orbit depends on the objectives of the sensor mission; orbit characteristics and different orbit types are explained below. A satellite must have a high speed to orbit at a certain distance from the Earth; the closer to the Earth the faster it has to move. A space station such as ISS has a mean orbital altitude of 400 km; it travels at roughly 27,000 km/h. The Moon being 384,400 km away can conveniently circle at only 3,700 km/h. At an altitude below 200 km, satellites already face traces of atmosphere, which causes rapid orbital decay. The higher the altitude the longer is the lifetime of the satellite. The majority of civil earth observing satellites orbits at an altitude in the range of 500 to 1000 km. Here we find the "big boys" such as Landsat-7 (2200 kg) and Envisat (8200 kg) but also the mini satellites of the Disaster Management Constellation (DMC). The DMC satellites have a weight of around 100 kg and were sent to space after the turn of the millennium by several countries around the world at relatively low-cost. They establish a network for disaster monitoring providing images in three or four spectral bands with a pixel size on the ground of 32 m or smaller.

Temporal resolution

Satellites have the advantage of continuity. Meteosat-9, *eg*, delivers a new image of the same area every 15 minutes and it does so every day for many years. The high 'temporal resolution' at low costs goes together with a low spatial resolution (pixel size on the ground of 3 x 3 km). Both the temporal and the spatial

resolution of satellite remote sensors are fixed. While aerial surveys were subject to restrictions in several countries, the access to satellite RS data is commonly easier. However, not every type of satellite RS images is universally available.

4.2.2 Aerial survey missions

Modern airborne sensor systems include a high-end GPS receiver and many also include an IMU. GPS is used for navigation and for coarse 'sensor positioning'. We need the coordinates of the exposure stations of a camera to relate points and features in the images to position on the ground. We apply differential GPS for more precise positioning. To this end we need a reference GPS station on the ground within some 30 km from the aircraft. Adding an Inertial Measuring Unit (IMU) has two advantages: the IMU readings can be used to improve the coordinates obtained by GPS (achieving a RMSE better than 0.1 m) and the IMU measures the attitude angles of the sensor (Figure 4.1). An IMU is an assemblage of gyros and accelerometers. It is a sophisticated, heavy, and expensive instrument, which was originally only used as part of an Inertial Navigation System (INS). Measuring continuously the position and attitude of the moving sensor allows us to relate the sensor recordings to position in the terrain in almost real-time. We call this *direct sensor orientation*. We need a GPS-IMU positioning and orientation system (POS) for line cameras and scanners; for frame cameras we can solve the georeferencing problem also indirectly (see Chapter 6).

Direct sensor orientation

Mission planning and execution is usually done by commercial survey companies or otherwise by large national mapping agencies or the military. Companies use professional software for flight planning and most likely one of the two integrated aircraft guidance and sensor management systems during missions (produced by APPLANIX or IGI). Pioneer work for computer controlled navigation and camera management was done at ITC when we still had an aerial photography and navigation department. The basics of planning aerial photography missions are explained in Chapter 9.

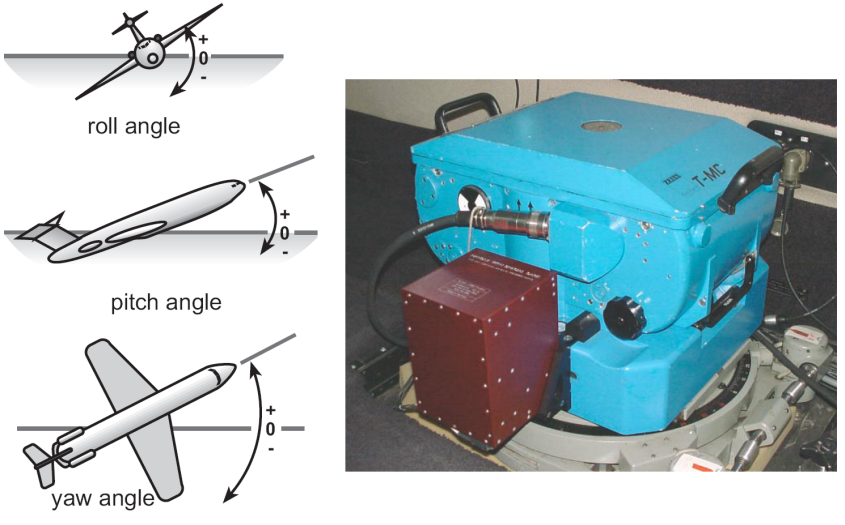


Figure 4.1: Attitude angles and IMU attached to an aerial camera Zeiss RMK-TOP (courtesy of IGI).

4.2.3 Satellite missions

Orbit parameters

The monitoring capabilities of a satellite sensor are to a large extent determined by the parameters of the satellite's orbit. An *orbit* is a circular or elliptical path described by the satellite in its movement round the Earth. Different types of orbits are required to achieve continuous monitoring (meteorology), global mapping (land cover mapping), or selective imaging (urban areas). For earth observation purposes, the following orbit characteristics are relevant.

- *Orbital altitude* is the distance (in km) from the satellite to the surface of the Earth. It influences to a large extent the area that can be viewed (*ie*, the 'spatial coverage') and the details that can be observed (*ie*, the 'spatial resolution'). In general, the higher the altitude the larger is the spatial coverage but the lower the spatial resolution.
- *Orbital inclination angle* is the angle (in degrees) between the orbital plane and the equatorial plane. The inclination angle of the orbit determines, together with the field of view (FOV) of the sensor, the latitudes up to which the Earth can be observed. If the inclination is 60°, then the satellite flies over the Earth between the latitudes 60° north and 60° south. If the satellite is in a low-earth orbit with an inclination of 60°, then it cannot observe parts of the Earth at latitudes above 60° north and below 60° south, which means it cannot be used for observations of the polar regions of the Earth.
- *Orbital period* is the time (in minutes) required to complete one full orbit. For instance, if a polar satellite orbits at 806 km mean altitude, then it has an orbital period of 101 minutes. The Moon has an orbital period of 27.3

days. The speed of the platform has implications on the type of images that can be acquired. A camera on a low-earth orbit satellite would need a very short exposure time to avoid motion blur due to the high speed. Short exposure time, however, requires high intensity of incident radiation, which is a problem in space because of atmospheric absorption. It may be obvious that the contradicting demands on high spatial resolution, no motion blur, high temporal resolution, long satellite lifetime and thus lower cost represent a serious challenge to satellite-sensor designers.

- *Repeat cycle* is the time (in days) between two successive identical orbits. The *revisit time* (*ie*, 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 combination to look to the side, or forward, or backward, not only vertically down. Many of the modern satellites have such a capability. We can make use of the pointing capability to reduce the time between successive observations of the same area, to image an area that is not covered by clouds at that moment, and to produce stereo images (see Section 4.5).

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

- *Polar orbit*. Polar orbit is an orbit with an inclination angle between 80° and 100°. An orbit having an inclination larger than 90° means that the satellite's motion is in the westward direction. Launching a satellite in eastward direction requires less energy, because of the eastward rotation of the Earth. Such a polar orbit enables observation of the whole globe, also near the poles. The satellite is typically placed in orbit at 600 km to 1000 km altitude.

- *Sun-synchronous orbit*. This is a near-polar orbit chosen in such a way that the satellite always passes overhead at the same time. Most sun-synchronous orbits cross the equator at mid-morning at around 10:30 hour local solar time. At that moment the Sun angle is low and the resultant shadows reveal terrain relief.
  In addition to day-time images, a sun-synchronous orbit also allows the satellite to record night-time images (thermal or radar) during the ascending phase of the orbit at the dark side of the Earth. Examples of polar orbiting, sun-synchronous satellites are Landsat, SPOT and IRS (see Section 4.6).

Orbit types

- *Geostationary orbit*. This refers to orbits where the satellite is placed above the equator (inclination angle: 0°) at an altitude of approximately 36,000 km. At this distance, the orbital period of the satellite is equal to the rotational period of the Earth, exactly one sidereal day. 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 (Figure 4.2). The geostationary satellites offer a continuous hemispherical view of almost half the Earth (45%), while the polar orbiters offer a higher spatial resolution.

RS images from satellites come with data on orbital parameters and other parameter values to facilitate georeferencing of the images. High resolution sensor systems such as Ikonos or QuickBird (see Section 4.6) use GPS receivers and star trackers as POS.

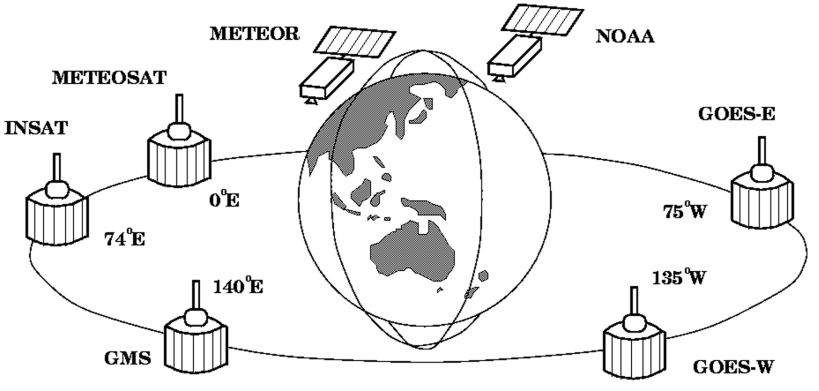


Figure 4.2: Meteorological observation by geostationary and polar satellites.

The data of spaceborne sensors need to be sent to the ground. Russia's SPIN-2 satellite with its KVR camera used film cartridges, which were dropped to a designated area on the Earth. Today the earth observing satellites 'downlink' the data. The acquired data are sent directly to a receiving station on the ground, or via a geostationary communication satellite. One of the current trends is that small receiving units, consisting of a small dish with a PC, are being developed for local reception of RS data.

4.2.4 Market figures

There is a large number of RS images from spaceborne sensors available from archives, also from satellites which are not operational anymore such as Lansat-1 MSS, SPIN, etc. In the following some figures are given on operational sensors, to give you some idea on market shares; these figures are only indicative and not exact.

Currently there are some 20 spaceborne optical sensor systems in operation when only counting those, whose data are most widely accessible. Compared to it the spaceborne radar systems are few - only three. Around 15 civil RS satellites are launched every year. You can consult ITC's Database of Satellites and Sensors for information about historical and current satellite remote sensors; there you will also find links to the data providers.

Aerial survey cameras have a long lifetime. An estimated 800 aerial film cameras are in service worldwide and new ones have still been bought in 2008. By mid 2008 about 300 large format digital aerial cameras have been sold (line and frame cameras together, by three manufactures). In 2008 it is nine digital aerial camera manufactures offering their products. The yearly turn over used to be 30 to 50 aerial survey cameras. In Europe an estimated 100 companies are operating in aerial surveying.

Around the turn of the millennium active sensing by laser scanning became a very booming business. There are some six major manufacturers of airborne

laser scanners, an estimated 250 systems in operation (mostly in Northern America and Europe), and more than 70 companies offering airborne laser scanning services. Compared to it the airborne radar market is small, with an estimated 20 systems in operation and only one commercial company providing airborne InSAR services (Intermap).

Airborne hyperspectral sensing and thermal scanning completes the spectrum of commercial primary data acquisition. Some 30 airborne imaging spectrometers are reportedly in operation and a similar amount of TIR scanners. NASA alone has/had four thermal scanners for various applications.

4.3 Cameras

A *digital camera* is an electro-optical remote sensor. In its simplest form, it consists of the camera body, a lens, a focal plane array of CCDs, and a storage device, but no mechanical component. The CCD array can either be an assembly of linear arrays or a matrix array (Figure 4.3). Accordingly we talk about line cameras and frame cameras. A small format frame camera has a single matrix chip and closely resembles a photographic camera. The chip (a) of the Figure 4.3 has 3 channels, one for each primary colour (red, green, blue); three elongated CCDs next to each other constitute one square "colour pixel". Each CCD has its colour filter right on top to only let the wanted band of incident light pass. The linear chip (b) of the Figure 4.3 also has three channels; three lines of square CCDs are assembled next to each other. A line camera is exclusively used on a moving platform, which can be a car, or an aircraft, or a spacecraft. The first satellite using a line camera was SPOT-1, which was launched in 1986. A line camera builds up a digital image of an area line by line (Figure 4.4). In older literature, therefore, it is also referred to as 'pushbroom scanner'; this opposed to a 'wiskbroom scanner' (see Section 4.4), which actually scans (across the track of the moving platform).

CCD

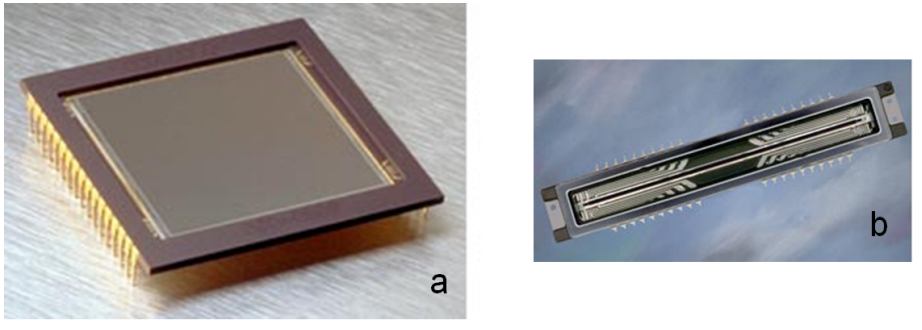


Figure 4.3: Two CCD chips: (a) matrix array Kodak KAF-16801, pixel size 9 μm; (b) linear array Kodak KLI-14403, pixel size 5 μm.

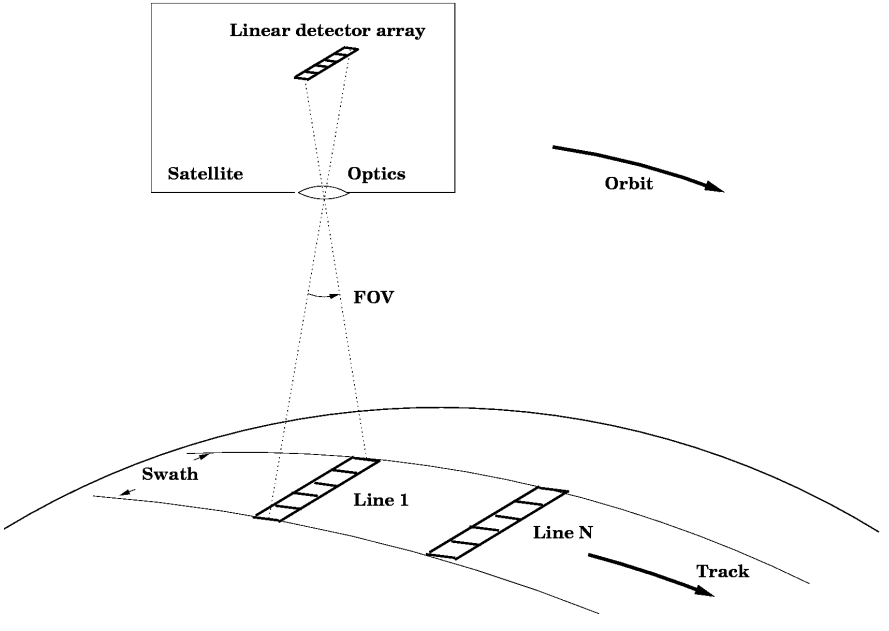


Figure 4.4: Principle of imaging by a line camera on a spacecraft ('pushbrooming').

4.3.1 Detector arrays

Cameras are used for sensing in the visible, NIR, and SWIR portion of the spectrum. We need different types of semiconductors for sensing in this range; the used semiconductors are all solid state detectors but made of different material for different spectral ranges. CCD is the most common type today for sensing in the visible to very near-IR range and it is made of silicon.

Spectral sensitivity

The spectral sensitivity of a sensor channel is commonly specified by a lower and an upper bound wavelength of the covered spectral band, *eg*, 0.48 to 0.70 μm for the SPOT-5 panchromatic channel. However, a detector such as a CCD is not equally sensitive to each monochromatic radiation within this band. The actual response of a CCD can be determined in the laboratory; an example of a resulting spectral response curve is shown in Figure 4.5. The lower and upper bound specification is usually taken at the wavelengths where the 50% response is achieved (see Figure 4.5). The DN produced by a detector results from averaging the spectral response of incident radiation. Figure 4.5 reveals that the DNs of AVHRR channel 1 are biased towards red whereas the brightness sensation of our eyes is dominated by yellow-green. The CCDs of a channel array do not have exactly the same sensitivity. It takes radiometric sensor calibration to determine the differences. CCDs, moreover, show varying degrees of degradation with time. Therefore, radiometric calibration needs to be done regularly. Knowing the detector's spectral sensitivity becomes relevant when we want to convert DNs to radiances (see Chapter 11).

When compared with photographic film we can say that normal CCDs have a much higher general sensitivity, thus they need less light. The reason is that

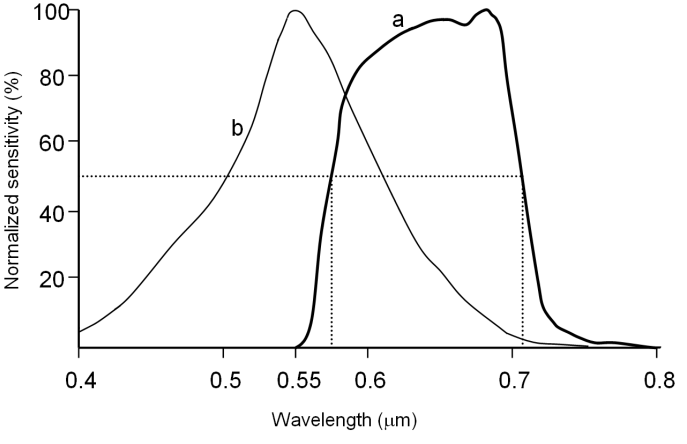


Figure 4.5: Normalized spectral response curve of (a) channel 1 of NOAA's AVHRR and (b) spectral sensitivity of the rods of the human eye.

they typically respond to 70% of the incident light whereas photographic film captures only about 2% of the incident light. They also offer a much better differentiation of intensity values in the very dark and the very bright parts of a scene.

We would like a CCD to have a wide dynamic range if we were interested in a high radiometric resolution. The *dynamic range* is the ratio between the maximum level and the minimum level of intensity that can be measured; it is also called the signal to noise ratio of the detector. The maximum is determined by the maximum charge capacity of the semiconductor cell. The minimum is determined by the noise level. Noise is unwanted collected charge, *eg*, stemming from unblocked IR or UV radiation for a CCD that should be sensitive to blue light. It only makes sense to record a DN with many bits if the semiconductor cell has a wide dynamic range. It is a manufacturer's concern to ensure a suf-

Dynamic range

ficient dynamic range to meet the user's requirement on radiometric resolution (expressed in bits). We can compute the effective radiometric resolution of a sensor if the manufacturer specifies both the number of bits and the dynamic range.

We had line cameras in space and frame cameras in our pocket before we had any digital airborne camera satisfying surveying quality . The main reasons have been the enormous quality of aerial film cameras and their operational maturity including the entire photogrammetric processing chain. Cameras on satellite platforms are exclusively line cameras, typically having a panchromatic channel and four more linear arrays (*eg*, for red, green, blue, NIR). ASTER has two panchromatic channels, one linear array looking vertically down (nadir view) and the second one looking backward; the two resulting image carpets can be used to generate stereo images. The first aerial line camera on the market was Leica's ADS40 (in 2000). It has three panchromatic detector arrays (forward, nadir, backward looking) and four multispectral ones (for RGB and NIR). One linear array has 12,000 CCDs.

Linear arrays

Current CCD technology allows producing very high quality linear arrays but not (yet) very large matrix arrays as we would need them for large format aerial cameras similar to the well proven film based survey cameras. The two market leaders in digital aerial frame cameras, ZI and Microsoft (former Vexcel), therefore, use several detector arrays for panchromatic imaging and compile a single large format image from the sub-frames by software. ZI's DMC, *eg*, has 13,500 x 7,500 CCDs per sub-frame. One of the advantages of the frame cameras is that the same photogrammetric software can be used as for photographs. At the moment there are about as many aerial line cameras as digital aerial frame cameras

Matrix arrays

on the market.

4.3.2 Optical system

Cameras use either lenses or telescopes to focus incident radiation to the focal plane where the CCD surfaces are. A lens of a simple hand-held camera is a piece of glass or plastic shaped to form an image by means of refraction. The lens cone of a survey camera contains a compound lens, which is a carefully designed and manufactured assembly of glass bodies (and thus very expensive; see Chapter 9 for more detail). The camera head of a digital aerial frame camera (such as Z/I's DMC and Vexcel's UltraCam) consists even of several such lenses to focus the light rays to the respective CCD arrays. However complicated a lens may be physically, geometrically the imaging through a lens is simple. The geometric model that a point of an object connects to its point in the image by a straight line and that all such lines pass through the centre of the lens (Figure 4.6) is a very close approximation of reality. We refer to the geometric imaging of a camera as 'central projection'.

Lens, focal length, scale

An important property of a lens is the 'focal length'. The focal length, f, determines together with the length of a CCD line the FOV of the camera. The focal length together with the 'flying height' determine the size of the ground resolution cell for a given pixel size, p. The *flying height*, H, is either the altitude of the aircraft above ground or the orbital altitude.

Field of view

$$GRC = p\frac{H}{f} \tag{4.1}$$

The focal length of Leica's ADS40 line camera is 63 mm. Considering a flying height of 2000 m we attain a ground resolution cell size in the across-track direction of 21 cm, if the airplane flies perfectly horizontally over flat terrain (see

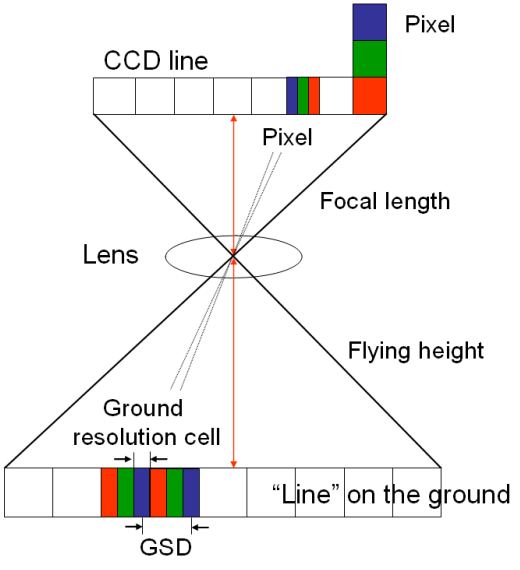


Figure 4.6: Pixel, ground resolution cell, ground sampling distance - for digital cameras.

Figure 4.6). You conclude correctly that the ADS40 has a CCD/pixel size of 6.5 µm. The ratio $\frac{H}{f}$ is referred to as *scale factor* of imaging.

Telescope

Spaceborne cameras do not have lens cones but telescopes. The telescopes of Ikonos and QuickBird consist of an assembly of concave and flat mirrors, thus achieving a spatial resolution, which is absolutely amazing when considering their flying height. The focal length equivalent of the Ikonos telescope is 10 m. Ikonos specifications state a ground resolution cell size of 80 cm for a panchromatic image at the nadir.

The size of a CCD determines the pixels size. A pixel projected onto the ground gives us the *ground resolution cell* (GRC) of the camera. The distance between the centres of two adjacent resolution cells of the same channel is called the *ground sampling distance* (GSD). Ideally the ground resolution cell size and the GSD are equal; the GSD then uniquely defines the spatial resolution of the sensor. This can be most easily achieved for panchromatic frame cameras. Note that the GSD is the same throughout an entire line if the terrain is flat and parallel to the focal plane (*eg*, in the case of a nadir view of horizontal terrain), see Figure 4.6. If a spaceborne line camera is pointed towards the left or the right of the orbit track (across-track off-nadir viewing), we obtain an oblique image. The scale of an oblique image changes throughout the image. In the case of oblique viewing, the Formula 4.1 does not apply anymore; the ground resolution cell size and the GSD increase with increasing distance from the nadir. Chapter 6 will explain how to deal with it.

Ground resolution cell

Ground sampling distance

Digital aerial cameras have several advantages as compared to film cameras, pertaining to both quality of images and economics. Digital aerial cameras commonly record in 5 spectral bands (panchromatic, RGB, NIR), therefore, we can obtain by one flight panchromatic stereo images, true colour images, and false colour images; with a film camera we would have to fly three times and develop three different types of film. The radiometric quality of CCDs is better than the one of photographic film material. Digital cameras allow for an all-digital workflow, thus it can be faster and cheaper. Digital cameras can acquire images with a large redundancy without additional costs on material and flying time; this favours automating information extraction. Finally, new geoinformation products can be generated as a result of the various extended camera features. In the next section multispectral scanners are introduced. Line cameras as compared to

Advantages of digital cameras

across-track scanners have the advantage of better geometry. Airborne line cameras and scanners require gyroscopically stabilized mounts to reduce effects of aircraft vibration and compensate rapid movements of the aircraft. Such a stabilized mount keeps a camera in an accurate level position so that it continuously points vertically downward. We want vertically taken images for mapping purposes because of the better geometry. Therefore, we also mount large format digital frame cameras and film cameras on stabilized platforms for applications requiring high quality images.

4.4 Scanners

4.4.1 Components

An *optical scanner* is an electro-optical remote sensor with a scanning device, which is in most cases a mechanical component. In its simplest form (*eg*, a thermal scanner operating in the 7 to 14 μm range), it consists of the sensor rack, a single detector with electronics, a mirror, optics for focusing, and a storage device (see Figure 4.7). A detector has a very narrow field of view, called the 'instantaneous field of view' (IFOV), of 2.5 milliradians or less. In order to image a large area we have scan the ground across the track while the aircraft or space craft is moving. The most commonly used scanning device is a moving mirror, which can be an oscillating mirror, or rotating mirror, or nutating mirror. An alternative, which is used for laser scanning, is fiber optics.

Detectors

Scanners are used for sensing in a broad spectral range, from light to TIR and beyond to microwave radiation. Photodiodes made of silicon are used for the visible and NIR bands. Cooled photon detectors (*eg*, using mercury cadmium telluride semiconductor material) are used for thermal scanners.

Beam splitters

Most scanners are multispectral scanners, thus sensing in several bands, often including TIR (such as NOAA's AVHRR). As such thermal scanners can be considered as being just a special type of multispectral scanners. A multispectral scanner has at least one detector per spectral band. Different from small format frame cameras where filters are used to separate wavelength bands, scanners and line cameras use a prism and/or a grating as 'beam splitter'. A *grating* is a dispersion device used for splitting up SWIR and TIR radiation. Also hyperspectral scanners use gratings. A *prism* can split up higher frequency radiation into red, green, blue, and NIR components. A simple RGB and NIR scanner

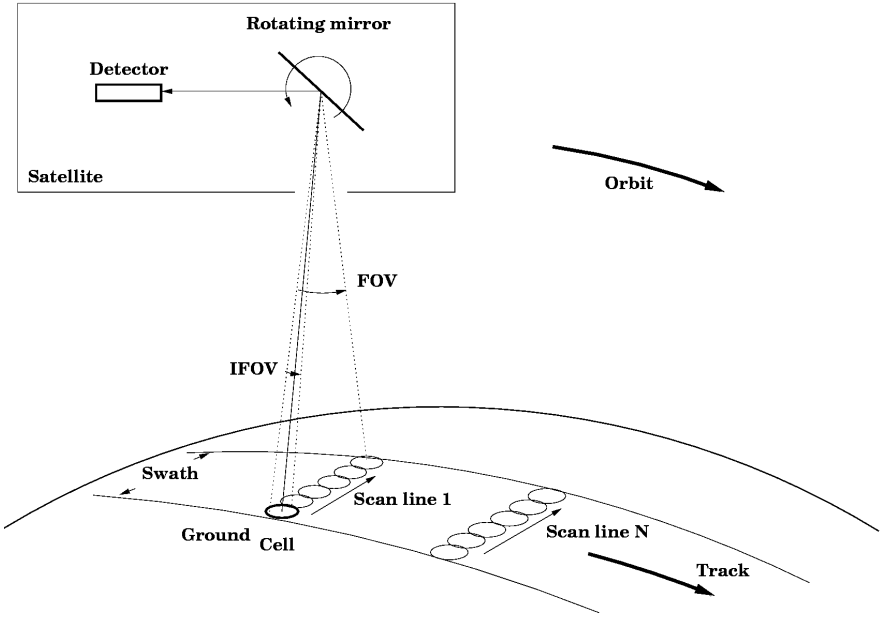


Figure 4.7: Principle of an across-track scanner.

produces in one sweep of the mirror one image line for each one of the four channels.

Instead of using only one detector per band, spaceborne scanners use several. The first civil spaceborne remote sensor, Landsat MSS (launched in 1972), used six per band (thus in total 24, see Figure 2.20). ASTER uses 10 detectors for each of its five TIR channels. One sweep of mirror of the ASTER thermal scanner produces thus 10 image lines for each of the five channels. If one channel should fail, only every 10th line of an image would be black. Chapter 11 treats the

correcting of an image for periodic 'line dropouts'.

4.4.2 Geometric aspects

At a particular instant, the detector of an across-track scanner observes an elliptical area on the ground, the ground resolution cell of the scanner. At the nadir the cell is circular of diameter D. D depends on the IFOV, β, of the detector and the flying height.

$$D = \beta \cdot H \tag{4.2}$$

A scanner with β = 2.5 mrad operated at H = 4000 m would have, therefore, a ground resolution of 10 m at the nadir. Towards the edge of a swath the ground resolution cell becomes elongated and bigger (Figure 4.8).

The width of the area, which is covered by one sweep of the mirror, the *swath width*, depends on the FOV of the scanner. AVHRR has a very wide FOV of 110°. Easy geometry was not a concern of the AVHRR design. Landast-7 has a FOV of only 15°, hence geometrically more homogeneous images result.

Reading out the detector is done at a fixed interval, the sampling interval. The sampling interval together with speed of the moving mirror determines the GSD. The GSD can be smaller than D. We talk about 'oversampling' if this is the case. The spatial resolution of the sensor is then not determined by the GSD but by the ground resolution cell size (which is larger or equal to D across the track).

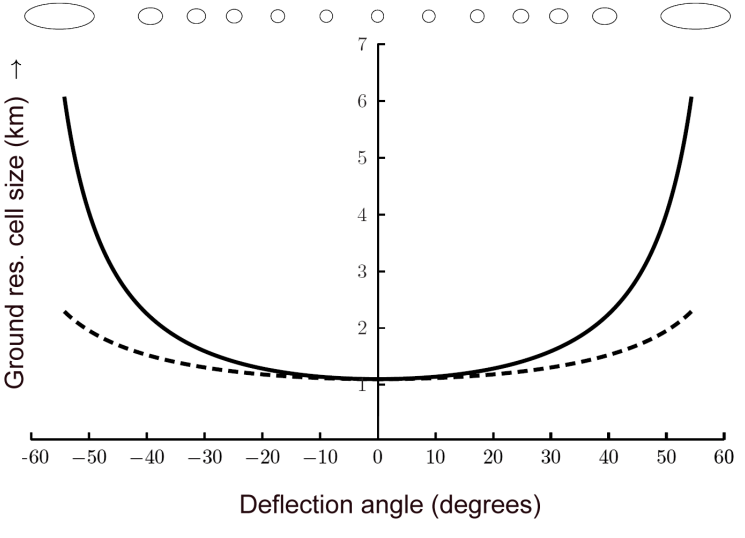


Figure 4.8: GRC of NOAA's AVHRR: at the nadir the cell diameter is 1.1 km, at the edge the ellipse stretches to 6.1 km by 2.3 km. The solid line shows the across-track resolution, the dashed line the along-track resolution. The ellipses at the top show the shape of the ground cells along a scanned line. NOAA processes the data ('resamples') to obtain a digital image with a pixel size on the ground of 1 by 1 km.

4.5 Stereoscopy

Stereoscopy as the science of producing three-dimensional (3D) visual models using two-dimensional (2D) images dates back to the 16^{th} century. The astronomer Kepler was presumably the first one to give a definition of stereoscopic viewing. A main reason for perceiving depth is that we have two eyes, which enable us to see a scene simultaneously from two viewpoints. The brain fuses the two 2D views into a three-dimensional impression. Judging which object is closer to us and which one is farther away with only one eye is only possible if we can use cues such as one object is partially obscured by the other one or it appears smaller although it is of the same size, etc. We can create the illusion of seeing three-dimensionally by taking two photographs or similar images and then displaying and viewing the pair simultaneously. Figure 4.9 illustrates the principle of stereoscopic viewing. The advantage of stereoscopic viewing above monoscopic viewing (looking at a single image) is that image interpretation is easier because we see the three-dimensional form of objects. Stereoscopy, moreover, has been the basis for 3D measuring by photogrammetry. Not any two images can be viewed stereoscopically, they must fulfill several conditions. The same holds for achieving 3D measurements; we need at least two images and they must meet the preconditions. The basic requirements for a 'stereo pair' are that the images of the same object or scene are taken from different positions, but not too far apart, and at a very similar scale. Different terms are used in stereoscopy with slightly different meanings. A pair of images that meets the conditions of stereoscopic vision may be referred to as stereo-image pair, stereoscopic pair of images, stereo images, or simply as stereo pair. A stereo pair arranged such (on the computer monitor, on the table, or in a device) that we can readily get a 3D visual impression may be called a stereograph, or stereogram). The 3D visual

impression is called the stereo model, or stereoscopic model. We need special image display techniques and stereoscopic viewing devices so that each eye sees only the image intended for it.

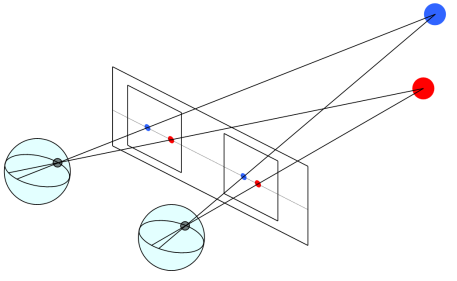


Figure 4.9: The principle of stereoscopy.

We have two options for obtaining a stereo pair with a spaceborne sensor: a) use across-track pointing to image the same area from two different tracks, or b) apply along-track forward or backward viewing in addition to nadir viewing. The advantage of 'in-track stereo' is that the two images are radiometrically very similar, because they are taken either at the same time or in quick succession; hence season, weather, scene illumination, and plant status are the same. In order to obtain a systematic coverage of an area by stereo images with an airborne frame camera we take strips of vertical photos/images, the images overlapping by at least 60% (see Chapter 9).

4.6 Overview of popular spaceborne sensors

This section characterizes major spaceborne earth observation systems and indicates their applications. The systems are listed starting with those of low spatial resolution and proceeding to higher resolution ones. As you will see, some satellites carry several sensors (and of different spatial and spectral resolution). The review of the multispectral systems is followed by a sketch of imaging spectrometers and the multi-instrument system Envisat, which includes an active microwave sensor. You can find a more detailed description of the sensors together with a specification of their main parameters in Appendix A; there you will also find a view on future developments.

Several space agencies/consortia launch new versions of their satellite-sensor combination in regular intervals (Meteosat, NOAA, Landsat, SPOT, etc). New versions are likely to be enhanced in one way or another, but remain similar in the choice of orbit, type of instruments, and spectral bands in order to guarantee data continuity. For example, in 2008 we have already Meteosat 8 & 9 operating and NOAA 17 & 18.

Meteosat/SEVIRI

The Meteosat satellites are the most remote ones today. They are geostationary and used by the World Meteorological Organization (WMO) together with another 30 to 40 geostationary and polar-orbiting satellites. The SEVIRI sensor records in the visible and several IR bands up to 13.4 μm. Main applications are rainfall forecasts and detecting thermal anomalies, which may be caused by forest fires, volcanoes, etc.

The NOAA satellites are polar-orbiting and carry several instruments.
The AVHRR scanner operates in the visible band and several IR bands up to 12.5 μm. 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, which can be used in climate monitoring, the study of El Niño, the detection of eddies to guide vessels to rich fishing grounds, etc. Cloud cover maps based on AVHRR data are used for rainfall estimates, which can be input into crop growing models. Monitoring shorelines, snow and ice, etc are examples of water related applications. A derived product of AVHRR data are the Normalized Difference Vegetation Index (NDVI) maps. These 'maps' give an indication about the quantity of biomass. 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. AVHRR data are appropriate to map and monitor regional land cover and to assess the energy balance of agricultural areas. A rather new application of Meteosat and NOAA data is to track cloud-free areas to optimize the data acquisition of high resolution spaceborne sensors.

NOAA/AVHRR

We first had Landsat-1 MSS. It was replaced in 1982 by Landsat-5 TM. In 2005 Landsat-5 TM operations had to be suspended for some time but the sensor is still operational today. The newer Landsat-7 ETM+ (launched in 1999) has not been fully operational since 2003. The Landsat satellites follow a sun-synchronous orbit. The ETM+ sensor records in eight bands with a spectral coverage from 0.45 to 12.5 μm. The data are used for many applications: land cover mapping, land use mapping, soil mapping, geological mapping, sea surface temperature mapping, monitoring deforestation, desertification, urbanization, etc. Landsat

Landsat/TM

data are preferred for land cover and land use mapping over SPOT multispectral data, because of the inclusion of middle infrared bands. Landsat 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, such as the crop temperature variability within irrigated areas due to differences in soil moisture. Appendix A includes a table, which lists typical applications for the various bands.

Terra

The Terra satellite with its five RS instruments has been designed to be highly synergistic, so that many interrelated problems can be solved. Terra also has been revolutionary in terms of providing data. Via internet it is easy to brows, purchase, and download data. The data are distributed worldwide at the costs of reproduction.

MODIS

MODIS observes the entire surface of the Earth every 1 to 2 days with an across-track scanning imaging radiometer. Its wide swath of more than 2300 km provides images of daylight reflected solar radiation and day-and-night thermal emissions over the entire globe. Its spatial resolution ranges from 250 m in the visible band to 1000 m in the thermal bands. MODIS has 36 narrow spectral bands, which are used to generate a variety of standard data products, ranging from atmospheric aerosol concentration, land and sea surface temperature, vegetation indices, thermal anomalies, snow and ice cover, to ocean chlorophyll and organic matter concentration products.

ASTER is a complex instrument, consisting of the following assemble: a nadir

viewing line camera with three channels in the visible and NIR range; a backward looking camera with a NIR channel for generating stereo images; a line camera with SWIR channels; a TIR scanner with 5 channels. The cameras and the scanner have independent pointing capability. The ASTER images are an excellent substitute for commercial Landsat images because they offer more bands, a better resolution, and a low price.

ASTER

SPOT-1 was revolutionary to topographic mapping because its line camera provided unprecedented spatial resolution of digital images from space combined with the capability of across-track viewing to obtain stereo images. The two HRG sensors of SPOT-5 have five channels with a spectral coverage from 0.48 to 1.75 µm. SPOT follows a sun-synchronous orbit like Landsat and Terra. Different from Landsat-7 ETM+, SPOT can provide stereo images. Different from Terra ASTER the stereo images are of much higher spatial resolution, *ie*, a GSD of 2.5 m in its supermode (panchromatic). SPOT - like Landsat and ASTER - serves a wide range of application but towards those requiring a higher spatial resolution as, *eg*, maritime surveillance. A prime application is medium scale topographic mapping. A peculiarity of SPOT-5 is that it carries a third camera, the panchromatic HRS instrument, which acquires stereo images almost simultaneously. The raw data of this camera are not sold; they are used exclusively to generate terrain surface elevation data at 30 m spacing and a global coverage.

SPOT

The Indian Space Research Organization has many operational RS missions and missions under development. Resourcesat-1, a sun-synchronous satellite, has followed up the famous IRS-1C and IRS-1D satellites. It carries three sensors (LISS4, LISS3, AWiFS), producing multi-band images in the visible and NIR

Resourcesat-1 and Cartosat

range. The satellite serves regional vegetation mapping and applications similar to those of Landsat. Cartosat-1 has been designed for mapping and in particular to produce terrain surface elevation data at 10 m spacing and orthoimages (see Chapter 6) of India. The panchromatic images, however, are available also for other regions. With a nominal GSD of 2.5 m it compares to SPOT/HRG, but the image quality used to be lower according to comparative investigations in 2007. On the other hand, Cartosat data cost less than SPOT-5 or Ikonos data.

Ikonos

Ikonos was the first entirely commercial "very high resolution satellite", following up the SPOT concept. Ikonos (the name comes from the Greek word for image) was also the first satellite not providing the RS community with raw data but only with different levels of geometrically processed data, with the aim to extend the buyer market to business and tourism. In 2000 panchromatic images with a GSD of 1 m were unprecedented. Different from SPOT, Ikonos and its conceptual successor QuickBird have the capability of pointing in any direction (across-track and along-track). One of the first tasks of Ikonos was acquiring images of all major USA cities. Ikonos data can be used for topographic mapping typically up to scale 1:10,000 and updating data held in topographic databases. Another interesting application is 'precision agriculture'; this is reflected in the multispectral band setting, which includes a NIR band. Regular updates of the "field situation" can help farmers to optimize the use of fertilizers and herbicides.

QuickBird (launched in 2001) and the recent WorldView are operated by DigitalGlobe, an USA company. Digital Globe distributes raw data and images processed to several levels of refinement. QuickBird images with a nominal GSD

QuickBird and WorldView

of 0.6 m for panchromatic images and 2.4 m for multi-band images represent in 2008 the highest resolution images from civil satellites. WorldView-1 - "the most agile satellite ever flown commercially" - only produces panchromatic images, with a nominal spatial resolution of 0.5 m and in-track stereo. WorldView-2 will have 8 multispectral bands with a GSD 1.8 m; it is planned for launch in September 2009. Typical applications are mapping, agricultural and urban planning, military surveillance, etc.

EO-1

Nasa's Earth Observing satellite EO-1 was launched in 2000 for a one-year technology demonstration mission. The line camera ALI mimicked Landsat ETM+ with the purpose to study cost reduction possibilities for a Landsat follow on. The hyperspectral sensor Hyperion has been the first imaging spectrometer operating from a spacecraft. It records in 220 bands in the range 0.4 to 2.5 µm. The third sensor on board has provided data for atmospheric corrections for multispectral sensors such Landsat ETM+. As of 2005 the EO-1 mission transitioned to lower cost and is now mainly an on-orbit test bed for hyperspectral research.

Proba/CHRIS

The European pioneer of spaceborne hyperspectral sensing was CHRIS on board of ESA's micro-satellite Proba. In 2006 CHRIS was still fully functional, recording in 63 bands of the visible to NIR range. Also intended for only a one-year technology demonstration mission it provided new biophysical and biochemical data as well as mineral information of land surfaces.

Envisat-1 is the most advanced satellite ever built by ESA. It carries in a sun-synchronous orbit 10 complementary, multidisciplinary sensors both active and

passive ones. The ASAR instrument represents the follow up of the famous ERS-1 and ERS-2 radar satellites. The medium resolution imaging spectrometer (MERIS) records in 15 bands of the visible to NIR range. Appendix A includes a table of the potential applications of MERIS data. In the Appendix you can also find information about the other sensors on board of Envisat-1.

Envisat-1

Trends in spaceborne RS

There have been trends towards higher spatial resolution (already 0.5 m GSD), higher spectral resolution (more than 100 bands, higher temporal resolution (global coverage within three days), and higher radiometric resolution (16 bits). Although most of these trends were driven by advancing technology, the trend towards higher spatial resolution was initiated by an act of politics, when the US president Clinton issued the US Land Remote Sensing Act of 1992. The act initiated a new 'Race to Space', this time about which private company would be the first to launch their high-resolution satellite. Today private companies spend more money on remote sensing than governments. Not only early industrialized countries are building and owning satellites. Korea, India, China, Brazil - just to name a few - have a strong ongoing remote sensing program, and we can expect that many more will enter into space.

4.7 Data selection criteria

4.7.1 Information requirements and constraints

For the selection of the appropriate data, it is necessary to fully understand the information requirements of a specific application. To this end you first have to analyze the spatial-temporal characteristics of the phenomena of your interest. For example, studying changing global weather patterns does not require spatially detailed information on terrain features. For identifying informal settlements, building details have to be recognized, thus we need data of high spatial resolution but not temporal differentiation. A specific spatial information requirement of several applications is modelling the vertical variation of the terrain surface; not all sensors can produce height or elevation data. In order to monitor slow and long-term processes, such as desertification or El Niño, we need data continuity during many years; ideally data characteristics that influence the needed information content - spatial, spectral, radiometric resolution - should not change over long periods.

The time for data acquisition may be constrained. It will be impossible to obtain flight permission for surveying a busy airport during daytime. We want to take aerial photographs at a time which allows for maximum interpretability. A low sun elevation causes long shadows of high buildings and trees, which obscure features we may be interested in. Therefore, we prefer to fly for aerial photographs around noon. In countries with a temperate climate, deciduous trees bear no leafs in the winter and early spring, therefore, there is a preference to execute photo-flights in these seasons to better see the ground. Farther north, however, the sun elevation may become too low during this period. Moreover, temperate climate zones have frequent cloud cover and occasionally suffer from

Sun elevation

Cloud cover

too strong winds or turbulence to permit good photo-flights. The consequence is that the 'aerial photo windows' are practically only a few days per year for countries at a latitude of Holland. Cloud cover is also a problem for spaceborne optical sensing - and very much so if the cloud cover is persistent as in areas of the tropical belt.

The information requirements need to be translated to data resolution and coverage taking into account data acquisition constraints. Spatial, spectral, radiometric, and temporal resolution have been explained in Sections 2.5, 4.2, and 4.3. By *spatial coverage* we mean the size of the area which can be viewed by the sensor at one moment (*eg*, the area covered by one photograph), or the area covered by the standard image size of a provider (one dimension of it is determined by the swath width; a standard SPOT image covers 60 km by 60 km). The *spectral coverage* is the total range of the spectrum that is covered by the channels of the sensor. Hyperspectral scanners usually have a coherent coverage (no gaps between the spectral bands) while multispectral sensors often have gaps in their spectral coverage (see Figure 2.20). The number of bands can be another criterion. A sole panchromatic camera has only one channel, a hyperspectral scanner as the EO-1/Hyperion yields a multi-band image with 220 bands. The *temporal coverage* is the span of time over which images are recorded and stored in image archives.

Spatial coverage

Spectral coverage

Temporal coverage

The *image size* is related to the spatial coverage and the spatial resolution. It is expressed as the number of rows (or lines) and number of columns (or samples) in one image. Remote sensing images contain thousands of rows and columns. The image file size in bytes can be calculated from the number of rows and columns,

Image size

the number of bands and the number of bits per pixel. 1 byte = 8 bits. For example, a four-band image may require as much as: 3000 rows $\times$ 3000 columns $\times$ 4 bands $\times$ 1 byte = 36 Mb of storage.

Geostationary weather satellites may generate much more data because of their almost continuous observation capability. For instance, every 15 minutes, Meteosat-8 generates a 3700 by 3700, 10 bits, 12-band image, which is: 3700 rows $\times$ 3700 columns $\times$ 12 bands $\times$ 1.25 bytes/pixel $\times$ 96 images/day = 20 Gb/day, which is 20 gigabyte of data per day. This means that Meteosat-8 alone generates several *terabyte* (10^{12} byte) per year. Some data archives are rapidly approaching the *petabyte* (10^{15} byte) limit.

4.7.2 Availability and cost

Once you have determined your data requirements, you have to investigate the data availability, accessibility, and costs. The availability depends on data that were already acquired and stored in archives, or data that need to be acquired at your request. The size and accessibility of image archives is growing at an ever faster rate. If up-to-date data are required, these need to be requested through an aerial survey company or from a remote sensing data provider.

Providing RS data is business, getting them for free is thus the exception. NOAA AVHRR is such an exception. Prices for satellite RS data have become negotiable. In general, low resolution data are cheaper per square km than data of high spatial resolution. Data from archives are cheaper than specially ordered data. High resolution satellite RS images are mostly offered at several processing levels (Ikonos, QuickBird, etc); in particular more refined geometric correction leads to a higher price tag. Low resolution RS images are mostly provided in a fixed size; for others a minimum order applies, but a polygonal area of interest can be specified. Especially for higher resolution data there can be considerable price differences between different parts of the world. Some data vendors have discount rates for educational institutes, or governmental organizations, or other target groups. The internet is the place to find out who currently provides what kind of satellite RS data, and under which conditions for a particular region.

Cost factors

The costs of aerial surveys depend on many factors (location, infrastructure, equipment, etc) and prices are often established in a bidding process. There are a few internationally operating air survey companies; certainly for them the ma-

jor cost factor is the deployment costs. In Europe the costs of aerial photographs are significantly lower than of images from high resolution satellite sensors.

Summary

This chapter has provided an introduction of platforms and passive sensors used for GDA. The two prime platforms are aircrafts and satellites. The key advantage of aircrafts is the possibility of targeted observation. Satellites can provide repeated and large area coverage. Most earth observation satellites circle in a 'sun-synchronous' orbit so that they pass overhead at the same time. The sensor-platform combination determines the characteristics of the obtained RS data, in particular the temporal resolution, the spatial resolution, and the spatial coverage.

We have distinguished cameras and scanners. Cameras have arrays of solid state detectors, a lens or telescope as optical system and no mechanical components. Line cameras have an assembly of only a few linear arrays. Frame cameras have one or several matrix arrays of detectors. Optical scanners have only a few semiconductor cells as detectors and a moving mirror to scan a scene across the track of the moving platform. Scanners and line cameras build up a 2D image of a scene line by line as the platform moves forward. Cameras "act on the high resolution playground", while scanners find their prime application in thermal infrared observations and hyperspectral sensing. Satellites travel very smoothly while aircrafts can suffer from turbulence. Therefore, airborne line cameras and scanners are mounted on stabilized platforms. Moreover, they also need a positioning and orientation system (POS) to facilitate relating the obtained images to position on the ground. 'Direct sensor orientation' by GPS and IMU is also increasingly used for aerial surveys with frame cameras.

The spectral sensitivity of a detector is not uniform within the spectral band covered and it changes with time. Repeated radiometric calibration is needed to enable adequate conversion of DNs to radiances. The spatial resolution depends on the characteristics of the detector assembly, the optical system, the 'flying height', and the pointing direction. What is often loosely specified as 'pixel size on the ground' is determined differently for cameras and scanners. For cameras the prime determining factors of spatial resolution are pixel size, focal length, and flying height. Spaceborne cameras and large format airborne cameras have for each spectral band a separate detector array; the 'pixel sixe on the ground' then corresponds to the 'ground sampling distance'. For scanners the prime influencing factors are the instantaneous field of view and the flying height as long as the total field of view is small.

The variety of platform-sensor combinations is as large as the range of applications of earth observing remote sensing. As a consequence, you are faced with the challenge to select the best data source for solving your particular mapping, monitoring, or modelling problem. The staring point should be analyzing your information requirements with respect to spatial-spectral-temporal characteristics of the phenomena of interest. In choosing an adequate data source you will have to cope with constraints such as data availability and data prices.

Additional reading

[13], [27], [17]; Chapters 7.3.2, 8.2, 9.

Questions

The following questions can help you to study Chapter 4.

1. Think of an application, define the spatial-spectral-temporal characteristics of interest and determine the type of remote sensing data required.
2. Which types of sensors are used in your discipline or field-of-interest?

The following are sample exam questions:

1. Explain the sensor-platform concept.
2. Mention two types of passive and two types of active sensors.
3. What is a typical application of a multi-band image from a spaceborne sensor, and what is a typical application of a very high spatial resolution image (from a spaceborne sensor)?
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 RS images.
7. Explain the principle of an across-track scanner.
8. Explain the principle of a line camera.

9. What does CCD stand for and what is it used for?
10. Consider a scanner with an IFOV of 2 mrad at a flying height of 5000 m. Calculate the diameter of the area observed at the nadir.
11. Explain the difference between IFOV and FOV.
12. Explain off-nadir viewing. What is an advantage?
13. Which ranges of spatial resolutions are encountered with today's multispectral and panchromatic sensors?

Chapter 5

Visualization and radiometric operations

5.1 Introduction

This chapter and the next one will explain the processing of raw remote sensing data and the basic principles of visualizing data. Producing images on a computer monitor or a printer always has been a prime concern of RS. We use images for inspecting raw data and for performing various data rectification and restoration tasks. Once data are corrected we convert them again to images and use these for information extraction by visual interpretation or to support digital image classification. Many RS applications make use of multispectral data; to visualize them we have to rely on colour. Section 5.2 explains how we perceive colour, which will help you to produce optimal images from multispectral data and properly interpret them.

We try to build remote sensors such that they faithfully image a scene. We are increasingly successful in dong so. Consider as example a vertical photograph or a nadir view of a high resolution spaceborne sensor - it closely resembles a map of the scene and an urban planner can readily recognize the objects of interest. Taking a closer look we know that RS images will geometrical be distorted as compared to a map. The degree and type of distortion depends on the sensor-platform type. Geometrically correcting RS images will be treated in Chapter 6. In Chapter 2 you have learned that remote sensors try to measure radiances but they record digital numbers, which have no direct physical meaning. The degree to which DNs directly correspond to radiances on the ground depends on many factors. Degradation with respect to what we would like to measure is caused by unfavourable scene illumination, atmospheric scattering and absorption, and detector response characteristics. The need to perform radiometric correction in

order to compensate any or all types of degradation depends on the intended application of the data. Our example urban planner or a topographic mapper does not need radiances of objects for object recognition in images. S/he, however, are likely to benefit from 'haze correction' and contrast enhancement to facilitate interpretation. Section 5.3, therefore, shortly treats radiometric corrections, only covering corrective measure of interest to a wider range of disciplines. You can find a detailed presentation on image restoration and atmospheric corrections in Chapter 11.

Elementary image processing techniques to improve the visual quality of an image - so that interpretation becomes easier - are introduced in Section 5.4. You may find image enhancement not only useful for earth observation but also for touching up your digital photos. Section 5.5 introduces image fusion; the further elaboration on image fusion in the remainder of this section belongs to the category of Capita Selecta of the book. Image fusion goes beyond simple radiometric operations; it involves radiometric, geometric, and spectral manipulations.

5.2 Visualization

5.2.1 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, *eg*, dark 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.

Tri-stimuli model

We experience light of different wavelengths 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 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 light (Figure 5.1). The signals sent to our brain by these cones give us colour-sensations. In addition to cones, we have *rods*, which sense brightness. The rods can operate with less light than the cones and do not contribute to colour vision. For this reason, objects appear less colourful in low light conditions.

Cones and rods

The knowledge of the tri-stimuli phenomenon led to the development of colour monitors. Colour television screens 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 of the cathode-ray 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 (*ie*, 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.

Colour monitors

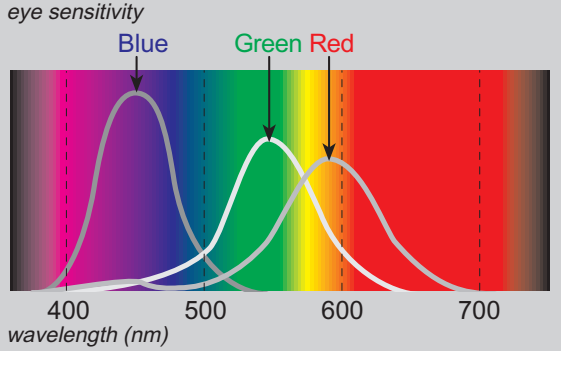


Figure 5.1: Visible range of the electromagnetic spectrum including the sensitivity curves of cones in the human eye.

Colour spaces

The tri-stimuli model of colour perception is generally accepted. It 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, which is based on the additive principle of colours.
2. Intensity-Hue-Saturation (IHS) space, which fits best our intuitive perception of colour.
3. Yellow-Magenta-Cyan (YMC) space, which is based on the subtractive principle of colours.

RGB

The RGB definition of colours is directly related to the way in which computer and television screens function. Three channels directly related to the red, green and blue dots are the 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 of the visible part of the spectrum. During the combination, the three colours are added. We see yellow when when green dots are illuminated in addition to red ones. This principle is called the *additive colour scheme*. Figure 5.2 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, so we experience 'white'.

Additive colour scheme

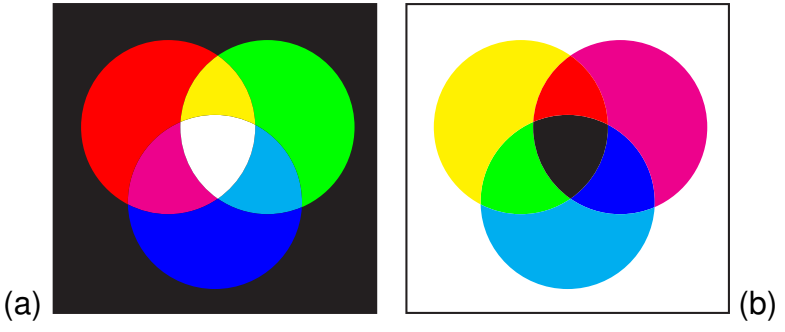


Figure 5.2: Comparison of the (a) additive and (b) subtractive colour schemes.

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 for red, green and blue, defining the so-called colour cube. Figure 5.3 shows the normalized colour cube, the maximum being set to 1.

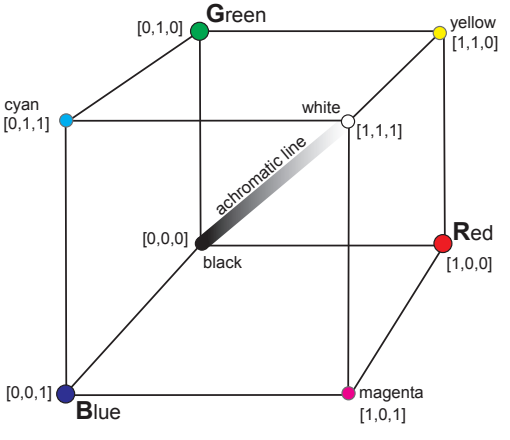


Figure 5.3: The RGB cube; note the red, green and blue corner points.

IHS

In daily speech we do not express colours using the RGB model. The IHS model more naturally reflects our sensation of colour. *Intensity* in the colour space describes whether a colour is dark or light and we use for intensity the value range 0 to 1. *Hue* refers to the names that we give to colours: red, green, yellow, orange, purple, etc. We quantify hue by degrees in the range 0 to 360. *Saturation* describes a colour in terms of purity and we express it as percentage. 'Vivid' and 'dull' are examples of common English language words used to describe colour of high and low saturation respectively. A neutral grey has zero saturation. As in the RGB system, again three values are sufficient to describe any colour.

Intensity

Hue

Saturation

Figure 5.4 illustrates the correspondence between the RGB and the IHS model. The IHS colour space cannot easily be transformed to RGB space because they are completely different. The cube in Figure 5.3 must be converted to a double cone; the inset of Figure 5.4 illustrates this. Although the mathematical model for this description is tricky, the description itself is 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 on the definitions does not exist. For further explanation you may want to consult the reference given at the end of the chapter. Important for image fusion (see Section 5.5) is the calculation of intensity values and luckly this is the simplest one. Be aware that the values in the actual RGB model range from 0 to 255, while in the IHS model the intensity ranges from 0 to 1. The formula for intensity is:

RGB to IHS

$$I = ((R + G + B)/3)/255. \tag{5.1}$$

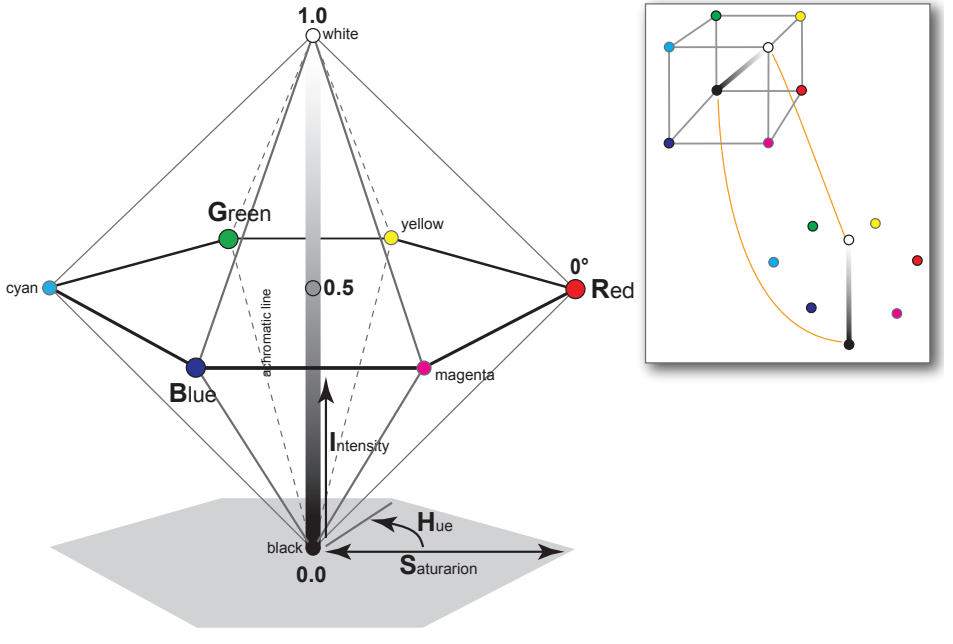


Figure 5.4: Relationship between RGB and IHS colour spaces.

An example: (R,G,B) = (150, 200, 100). I = ((150+200+100)/3)/255 = 0.59.

YMC

Whereas RGB is used for computer and TV screens, the YMC colour description is used in colour definition for hardcopy material, such as printed pictures and 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 of it is filtered out by the ink layers and the remainder is reflected from the underlying paper (see Figure 5.2).

Subtractive colour scheme

5.2.2 Image display

We normally display a digital image using a grey scale. A digital image can be raw data as obtained by a panchromatic camera, or data as obtained by scanning a B&W photograph, or a single band of a multi-band image. Standard computer monitors support for image display 8 bits per dot. Thus, if we have sensor recordings of 8 bits, each DN will correspond to exactly one grey value. A pixel having the value zero will be shown as black, a pixel having the value 255 as white. Any DN in between becomes some shade of grey. The one to one correspondence DN - grey value used to be the standard, so we still often use 'grey value' as synonym for DN. A colour monitor has three input channels, so we have to feed each of them with the same DN to obtain a 'grey scale image' (Figure 5.5).

Grey scale

Figure 5.5: Single-band and three-band image display using the Red, Green, and Blue input channel of a monitor.

An alternative way to display single-band data is to use a colour scale to obtain a 'pseudo-colour' image. We can assign colours (ranging from blue via cyan,

green and yellow to red) to different portions of the DN range from 0 to 255 (see Figure 5.5). 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 versus warm with blue versus red is more intuitive than with dark versus bright.

Pseudo-colour

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 intended use of the image. Figure 5.5 indicates how we obtain a 'false colour composite'.

Colour composites

Sometimes a *true colour composite* is made, where the RGB channels relate to the red, green and blue wavelength bands of a camera or multispectral scanner. An other popular choice is to link RGB to the near-infrared, red and green bands respectively to obtain a *false colour composite* (see Figure 5.6). The results look similar to colour infrared photographs. As explained in Chapter 9, the three layers of a colour IR 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, but their infrared reflection is even higher. Therefore, vegetation displays in a false colour composite as a combination of some blue and a lot of red, resulting in a reddish tint of purple.

True colour

False colour

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-5 TM bands 5–4–3 or 4–5–3 to RGB. Combinations that display NIR as green show vegetation in a green colour and are, therefore, often called *pseudo-natural colour composites* (see Figure 5.6). It must be mentioned that there is no common consensus on the naming of certain composites ('true' may also be referred to as 'natural', 'false colour' may also be used for other band combinations than green, red and NIR, etc). Once you have become familiar with the additive mixing of the red, green and blue primaries, you can intuitively relate the colours - which you perceive on the computer monitor - to the digital numbers of the three input bands, thereby gaining qualitative insight in the spectral properties of an imaged scene.

Pseudo-natural colour

To obtain a 3D visual model from a stereo-image pair on a computer screen, we must combine the images to a stereograph (Figure 5.7) and then use a device which helps us to view the left image with the left eye and the right image with the right eye. There are various technical solutions to this problem. One of them is the 'anaglyph' method. The left image is displayed in red, the right one in cyan and the two images are superimposed. For viewing we need spectacles with a red glass for the left eye and a cyan glass for the right eye. High end digital photogrammetric systems use polarization instead of colour coding. Polarizing spectacles make the images visible to the respective eyes. The advantage of using polarization is that we can achieve a more accurate colour viewing than with an anaglyph stereograph of colour images, or superimpose the results of measuring in a panchromatic stereo model in any colour. A yet other approach is to use a 'split screen' display and a stereoscope in front of the monitor. A stereoscope is a device consisting of a binocular and two mirrors so that we can put two images next to each other to achieve stereoscopic vision. We can also

Anaglyph stereograph

Stereoscope

use stereoscopes to view paper prints of stereo photographs.

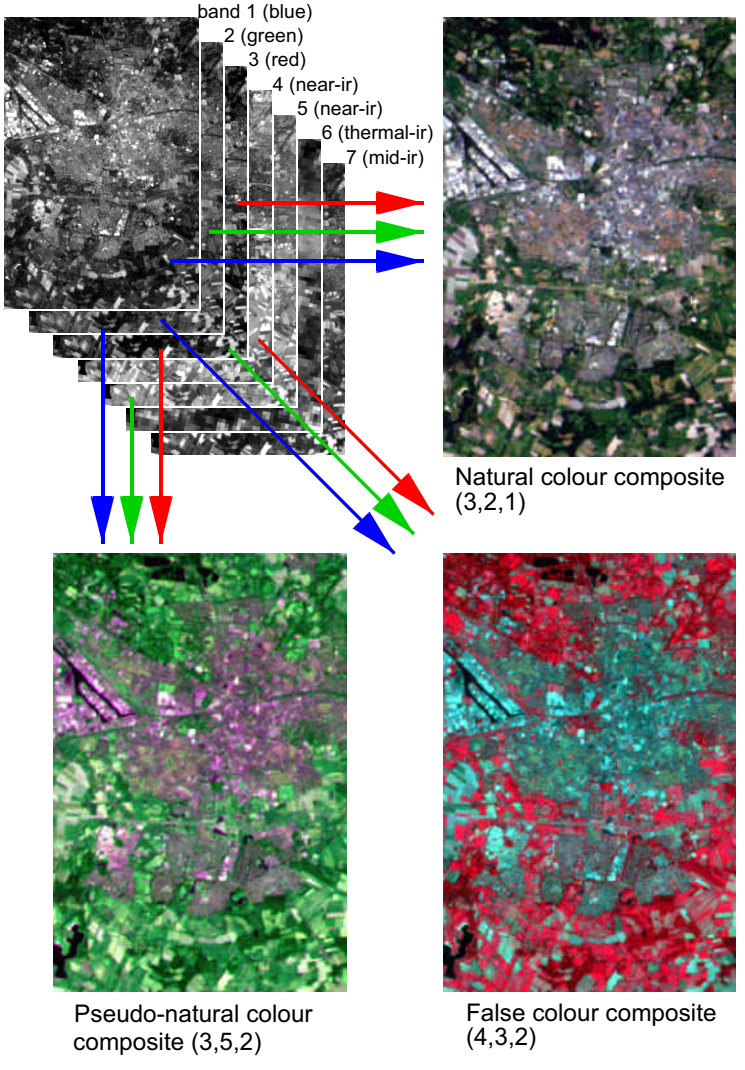


Figure 5.6: Landsat-5 TM false colour composite of Enschede and surrounding. Three different colour composites are shown: true colour, pseudo-natural colour and false colour composite.

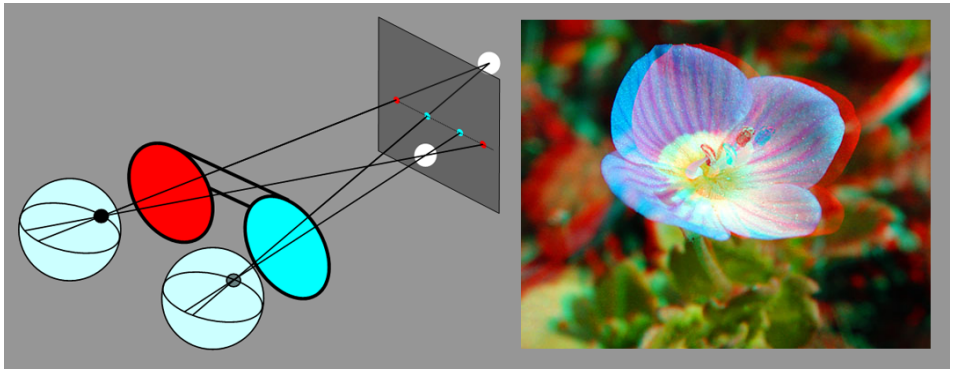


Figure 5.7: Anaglyph principle and stereograph.

5.3 Radiometric corrections

Under the heading of radiometric corrections, we can group various techniques, which improve radiometrically degraded raw RS data and serve different interest. Radiometrically correcting data should make them fit better for information extraction. Techniques of modifying recorded DNs serve any of three main purposes as outlined below.

- Correcting data for imperfections of the sensor. The detector cells of a camera have all a slightly different response. We can determine the differences by radiometric calibration and apply later accordingly radiometric correction to the recordings of the camera. Scanners often use several detectors per channel instead of only one in order to scan a wider strip on the ground with one mirror sweep. Again, the detectors will have (slightly) different radiometric response with the consequence that the resulting image may show stripes. A 'destriping' correction will normalize the lines relatively. Moreover, a worse case is that one detector fails. We then obtain an image where, *eg*, every 10th line is black. A 'line drop' correction will cosmetically fix the data. Another detector problem is random noise, which degrades radiometric information content and makes a RS image appear as if salt and pepper was added to the scene. Correcting the above mentioned disturbances is fairly simple and explained in Chapter 11. There can be other degradations caused by the sensor-platform system that are not so easily corrected. An example is compensating for image motion blur, which relies on a mathematical complex technique. We got used to referring to theses types of radiometric corrections as 'image restoration'. Luckily image restoration of new sensor data is commonly done by the

data providers. You may only have to apply techniques such as destriping and dropped line correction when dealing with old data, *eg*, from Landsat MSS. Image restoration should be applied prior to further corrections and enhancements.

Image restoration

- Correcting data for scene peculiarities and atmospheric disturbances. One of the "scene peculiarities" is how the scene is illuminated. Consider an area at an appreciable latitude like Holland. The illumination of the area will be quite different in winter and in summer (overall brightness, shadows, etc) because of different sun elevation. Normalizing images taken in different seasons is briefly outlined below. An atmospheric degradation effect, which is already disturbing when extracting information from one RS image, is atmospheric scattering. Skyradiance at the detector causes haze in the image and reduces contrast. Haze correction is briefly described below. Converting DNs to radiances on the ground (Chapter 11) becomes relevant if we want to compare RS data with ground measurements, or if we want to compare data acquired at different times by different sensors for change detection.

  Scene normalization

  Atmospheric correction

- Enhancing images so that they are better suited for visual interpretation. Image enhancement techniques are introduced in a separate section because they can be taken a step further, namely to 'low-level image processing' for computer vision.

  Image enhancement

5.3.1 Sun elevation correction

Seasonal illumination differences will be disturbing if we want to analyze sequences of images of the same area but taken at different times, or if we would like to make mosaics of such images. We can apply a simple sun elevation correction if the images stem from the same sensor. The trick is to normalize the images such as if they were taken with the sun at the zenith. We can achieve it by dividing every pixel value of an image by the sine of the sun elevation angle at the time of data acquisition. The sun elevation angle is usually given in the metadata file, which is supplied with an image. Obviously this is an approximate correction, which does not take into account the effect of elevation and height difference in the scene nor atmospheric effects. If the sun elevation angle is not available, we can use measurements of stable land cover as reference for this correction. DNs of stable land cover are then considered to become equal.

Normalization by sine

5.3.2 Haze correction

Formula 2.4 has shown that atmospheric scattering adds a 'skyradiance'. Haze correction aims at removing this effect from raw data. Doing so can be beneficial to many applications of spaceborne RS. In Chapter 2 you have also learned that scattering depends on wavelength; Rayleigh scattering will hardly affect the recordings in the red spectral band while the DNs in the blue band may become significantly larger. Reducing haze, therefore, must be done independently for each band of a RS image. How much to subtract from every DN of a particular band? We can find out if the scene is favourable and contains areas that should have zero reflectance (a spectral band specific black-body). Deep clear water, *eg*, should yield zero pixel values in an IR band. If not, we attribute the minimum value found for "water pixels" to skyradiance and subtract this value from all DNs in this band. The alternative is less reliable, *ie*, is to look at the histogram (see Section 5.4.1) of the band and simply take the smallest DN found there as the haze correction constant.

Subtraction per band

5.4 Elementary image processing

The two categories of elementary image processing to enhance an image are histogram operations and filtering. In order to increase the visual distinction between features, histogram operations aim at global contrast enhancement, while filter operations aim at enhancing brightness edges and suppressing unwanted image detail. Histogram operations look at pixels without considering where they are in the image and assign a new value to a pixel by making use of a look-up table, which is set up from image statistics. Filtering is a "local operation", computing the new value for a pixel based on the values of the pixels in the local neighbourhood. Figure 5.8 shows the effect of contrast enhancement and edge enhancement as applied to the same input image. Section 5.4.1 first explains the notion of histogram.

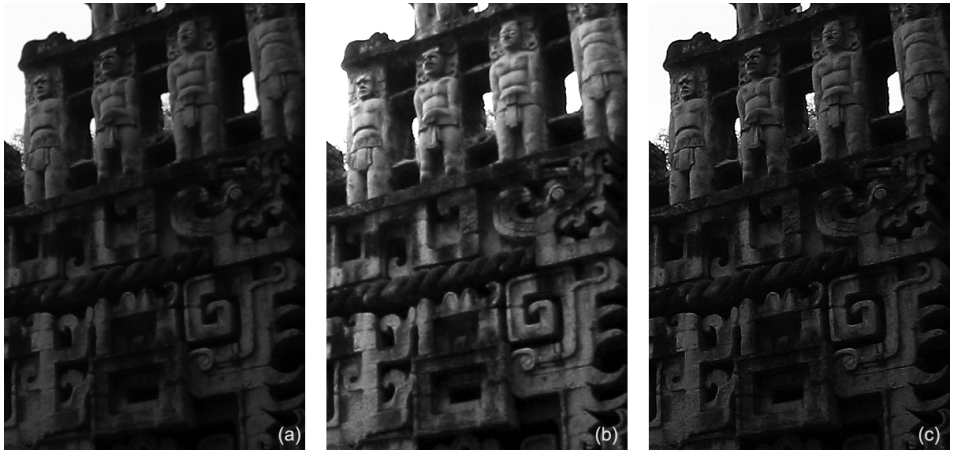


Figure 5.8: Original (a), contrast enhanced (b), edge enhanced (c) image.

5.4.1 Histograms

The radiometric properties of a digital image are revealed by its histogram. By changing the histogram we change the visual quality of the image. The *histogram* describes the distribution of the pixel values of a digital image. These DNs are in the range from 0 to 255. A histogram shows the number of pixels for each value in this range, *ie*, the frequency distribution of the DNs. Histogram data can be represented either in tabular form or graphically. The tabular representation (Table 5.1) usually shows five columns. From left to right these are:

Frequency distribution of DNs

- 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 the DN
- CumPerc: cumulative frequency as a percentage of the total number of pixels.

Figure 5.9 shows a plot of the columns 3 and 5 of Table 5.1 against column 1. More commonly the histogram is displayed as bar graph rather than as line graph. The graphical representation of column 5 can readily be used to find the '1% value' and the '99% value'. The 1% value is the DN, below which only 1% of all the values are found. Correspondingly there are only 1% of all the DNs of the image larger than the 99% value. The 1% and 99% values are often used

1% and 99% cut-off

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 5.1: Example histogram in tabular format (only some sections).

in histogram operations as cut-off value, considering very small and very large DNs as noise rather than as signal.

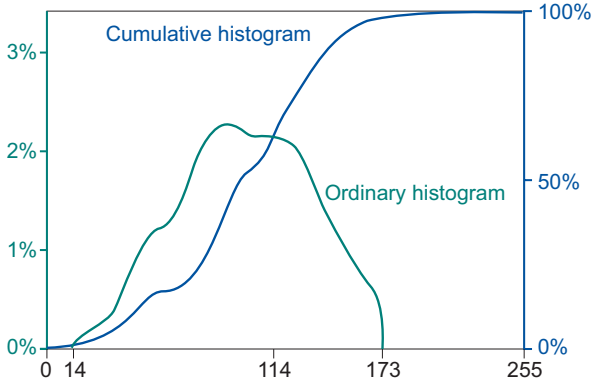


Figure 5.9: Standard histogram and cumulative histogram corresponding with Table 5.1.

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

Table 5.2: Summary statistics for the example histogram given above.

A histogram can be summarized by descriptive statistics: mean, standard deviation, minimum and maximum, as well as the 1% and 99% values (Table 5.2). The mean is the average of all the DNs of the image; it often does not coincide with the DN that appears most frequently (see Table 5.1 and Table 5.2). The standard deviation indicates the spread of the DNs around the mean.

A narrow histogram (thus, a small standard deviation) represents an image of low contrast, because all the DNs are very similar and mapped to only a few

grey values. Figure 5.11(a) displays the histogram of the image, which is shown in Figure 5.8(a). You can notice a peak at the upper end (DN larger than 247), while most of DNs are smaller than 110. The peak for the white pixels stems from the sky. All other pixels are dark greyish, this narrow histogram part characterizes the poor contrast of the image (of a Maya monument in Mexico). Remote sensors commonly use detectors with a wide dynamic range, so that they can sense under very different illumination or emission conditions. Very different illumination conditions, however, are not always given within one particular scene. In practice we often obtain, therefore, RS images that do not exploit the full range for DNs. A simple technique to then achieve better visual quality is to enhance the contrast by a 'grey scale transformation' which yields a histogram stretched over the entire grey range of the computer monitor (Figure 5.11(c)).

Poor contrast

5.4.2 Histogram operations

We may have two different purposes for a contrast enhancement, but in both cases we want to merely promote the visual interpretability of an image. The first purpose is "temporary enhancement" where we do not want to change the original data but only get a better picture on the monitor so that we can do a certain interpretation task. An example is image display for geometric correction (Chapter 6). The other purpose would be to generate new data, which have higher visual quality. We want to do so when an image product will be the output of our RS endeavour. An example is orthophoto production and image mapping (see Chapter 6). Below two techniques of contrast enhancement are described, 'linear contrast stretch' and 'histogram equalization' (occasionally also called histogram stretch. Both are 'grey scale transformations', which convert input DNs (of raw data) to new DNs (of a more appealing image) by a user defined 'transfer function'.

Linear contrast stretch

Linear contrast stretch is a simple grey scale transformation where the lowest input DN of interest becomes zero and the highest DN of interest becomes 255 (Figure 5.10). The monitor will display zero as black and 255 as white. As lowest and highest input DN we often take the 1% and 99% values. The functional relationship between the input DNs and output pixel values is a linear one as shown in the Figure 5.11(a). The function shown in the first row of the Figure 5.11 (against the background of the histogram of the input image) is called the *transfer function*. Many image processing software packages allow to graphically manipulating the transfer function so that we can obtain an image appearance of our liking. The implementation of the transformation can be done using

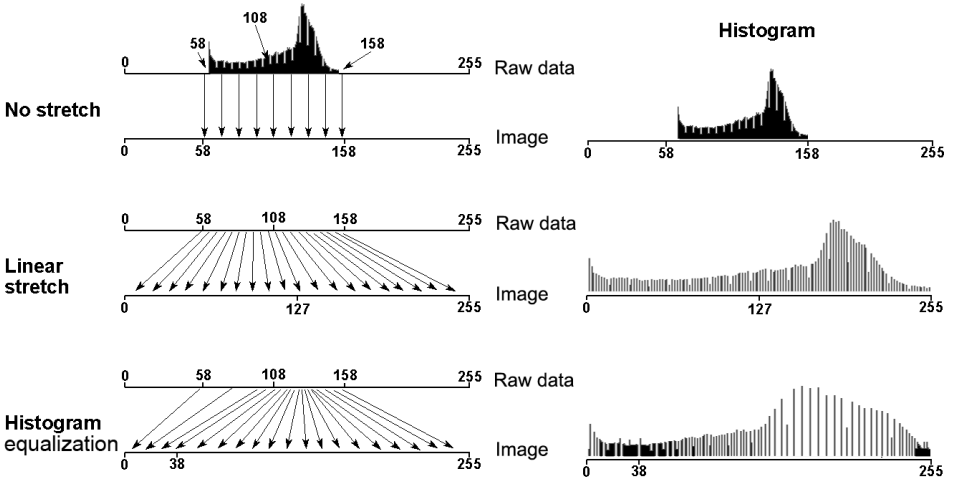


Figure 5.10: Linear contrast stretch versus histogram equalization.

a look-up table as indicated in Figure 5.10.

The transfer function can be chosen in a number of ways. We can use the linear grey scale transformation to correct for haze and also for other purposes than contrast stretching. For instance, to 'invert an image' (convert a negative image to a positive one or vice versa), or to produce a binary image (as the simplest technique of image segmentation). Inverting an image is relevant, *eg*, when having scanned a photographic negative film, see the last column of Figure 5.11.

Transfer function

Linear stretch is a straight forward method of contrast enhancement, giving fair results when the input image has a narrow histogram but close to uniform distribution. The histogram of our example image (Figure 5.8(a)) is asymmetric,

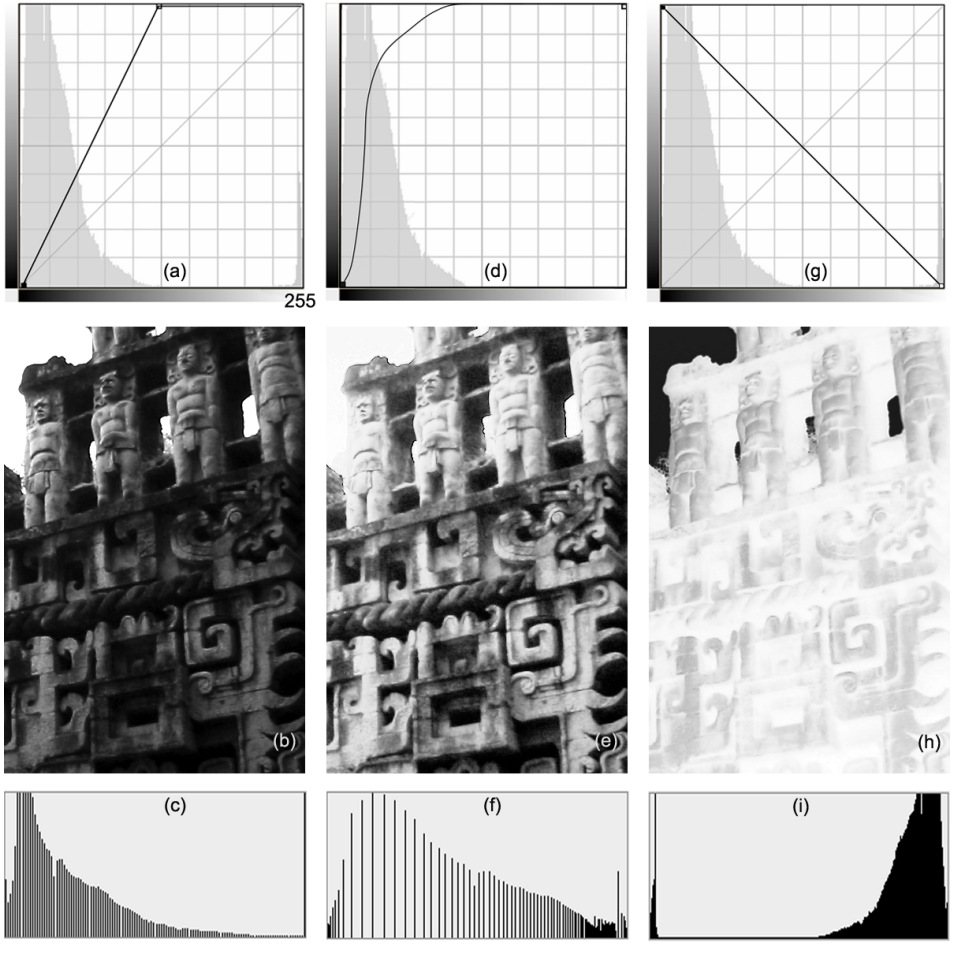


Figure 5.11: Effect of linear grey scale transformations (b,c & h,i) and histogram equalization (e,f).

Histogram equalization

with all DNs in a small range at the lower end, when not considering the irrelevant sky pixels (Figure 5.11(d)). Stretching this small range at the expense of compressing the range with only few values (in our example the range of high brightness) will lead to seeing more detail in the dark parts (see Figure 5.11(e)). As the name suggest, *histogram equalization* aims at achieving a more uniform histogram (see Figure 5.11(f)). Histogram equalization is a non-linear transformation; several image processing packages offer it as a default function. The idea can be generalized to achieve any desired target histogram.

It is important to note that contrast enhancement (by linear stretch or histogram equalization) merely amplifies small differences in the data so that we more easily can visually differentiate features, but it does not increase the information content of the data. Histogram operations should be based on analyzing the shape and extent of the histogram of the raw data and having an understanding of what is relevant for the wanted interpretation. A decrease of information content can otherwise readily result.

5.4.3 Filter operations

A further step in producing optimal images for interpretation is to apply filtering. Filtering is usually carried out on a single band. Filters can be used for image enhancement, for example, to reduce noise ("smooth an image") or to sharpen a blurred image. Filter operations are also used to extract features from images, *eg*, edges and lines, for automatically recognizing patterns and detecting objects.

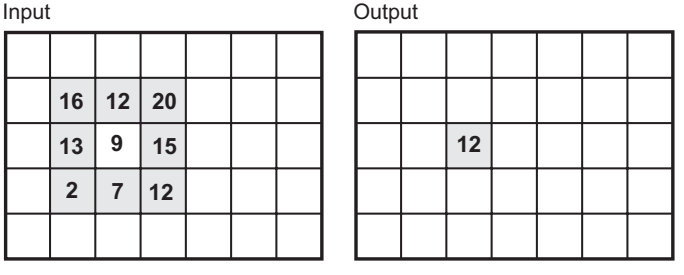


Figure 5.12: Input and output of a filter operation: the neighbourhood in the original image and the filter kernel determine the value of the output. In this situation a smoothing filter was applied.

A large group of filters are the linear filters. They calculate the new value of a pixel as a linear combination of the given values at the pixel under consideration and its neighbours. A simple example of a smoothing filter is to compute the average of the pixel values of a 3 by 3 neighbourhood and assign it as new value to the central pixel (Figure 5.12). We can conveniently define such a linear filter through a 'kernel'. Table 5.3 shows the kernel of the smoothing filter applied in the example of Figure 5.12. The kernel specifies the size of the neighbourhood that is considered (3 by 3, or 5 by 5, or 3 by 1, etc) and the coefficients for the linear combination. Image processing software allows us to select a kernel from

Linear filter

a list and/or define our own kernel. The sum of the coefficients for a smoothing filter should be one, otherwise an unwanted scaling of grey values will result. The filtering software will calculate the 'gain':

$$gain = \frac{1}{\Sigma k_i} \tag{5.2}$$

and multiply the kernel values with it. The following two subsections give examples of kernels and their gain. Since there is only one way of using the kernel of a linear filter, the kernel completely defines the filter. The actual filter operation is to "move the kernel over the input image" line by line, calculating for every pixel a local 'average' of pixel values. A linear filter, therefore, is also called a 'moving average' (average is to be understood as any linear combination of input values, not just as arithmetic mean).

Moving average

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.

Noise reduction

Consider the kernel shown in Table 5.3 in which all values 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 an image will become blurred (smoothed). When dealing with the speckle effect in radar images the result of applying this filter is reduced speckle.

1	1	1
1	1	1
1	1	1

Table 5.3: 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 instead of an arithmetic mean. 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 5.4.

1	2	1
2	4	2
1	2	1

Table 5.4: Filter kernel for weighted smoothing.

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

Table 5.5: Filter kernel used for edge enhancement.

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, etc. 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 elements. An example of an edge enhancement filter is given in Table 5.5.

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 5.13.

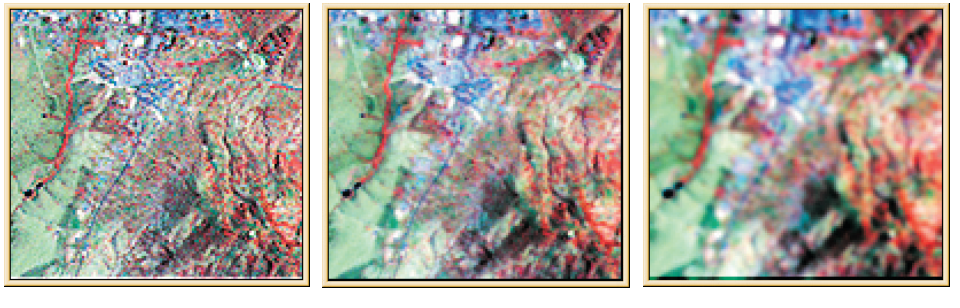


Figure 5.13: Original image (middle), edge enhanced image (left) and smoothed image (right).

5.5 Image fusion

Digital aerial cameras and many high resolution spaceborne cameras record simultaneously multispectral and panchromatic data. Most of these cameras use for the panchromatic channel a spatial resolution that is about a factor four higher than the resolution of the RGB and NIR channels. This corresponds to the resolution capabilities of the cones and rods of our eyes. Colour images are easier to interpret than grey scale images. Higher resolution images are easier to interpret than lower resolution ones. Why not to combine the advantages of both? It can be accomplished by a technique called pan-sharpening. *Pan-sharpening* is an image fusion technique, which combines three images from the multispectral channels with the higher resolution panchromatic image to a colour composite. Pan-sharpened true colour or pan-sharpened false colour images have become a standard product of many data providers. Pan-sharpening is only one of the many image fusions you can learn to apply yourself. Image fusion is about combining RS data from different sources with the purpose to obtain one image on the computer screen, which promotes interpretation. Other examples of exploiting complementary characteristics than (i) fusing high spatial resolution with colour are combining: (ii) the textural properties of a synthetic aperture radar image with the multispectral properties of an optical data set, (iii) various terrain properties derived from a digital terrain relief model (see Chapter 6) with a single-band image, (iv) gridded airborne geophysical measurements with relief shades computed from a digital terrain relief model or satellite data of higher spatial resolution, and (v) images taken at different times for change detection.

Pan-sharpening

Fusion examples

The advanced reading section below explains the basic principles of image fu-

sion.

The principle of colour coding of multispectral data as explained in Section 5.2.2 can also be applied to images stemming from different sensors. The aim of image fusion is to make optimal use of sensor specifics and to combine complementary information. The processing technique underlying all image fusion methods is based on applying a mathematical function on the co-registered pixels of the merged image set, yielding a single image optimized for visual interpretation. Without considering pre-processing steps such as atmospheric corrections and the removal of striping and other noise artifacts (see Chapter 11), the image processing procedures employed in image fusion can be subdivided into four steps:

1. Geometric co-registration of the images to be fused
2. Enhancement of individual image channels to be fused
3. Application of an arithmetic or other mathematical function to the co-registered pixels of the multi-band image
4. Visualization of the fused data on a computer monitor.

The first step refers to the geometric registration procedures that will be explained in Chapter 6. Whether we can employ a two-dimensional approach (geometric corrections that ignore relief displacement) is dependent on the images to be fused. It is crucial to register the images to sub-pixel accuracy, because mismatches between images that are registered at a lower accuracy become obvious as distracting coloured artifacts in the image fusion result. The selection of the

final spatial resolution for the fused image is included in this first step. It is a subjective choice that depends on the type of data involved and on the interpretation problem to which image fusion is applied. In most applications of image fusion, however, a pixel size is chosen that is equal or similar to that of the image with the highest spatial resolution. This allows us to exploit image details in the fused product that may not be apparent in the lower spatial resolution images.

The second step is often used to ensure that features of interest are optimally enhanced in the image fusion result. This step usually employs common enhancement techniques applicable to single-band images, such as contrast enhancement (Section 5.4.2) and filtering (Section 5.4.3). It also requires sufficient knowledge on how the properties of an image acquired by a particular sensor system contribute to solving the interpretation problem.

One commonly used method to fuse co-registered images from multiple sources is to skip the third step and directly display the georeferenced images as a colour composite. Although this is the simplest method to fuse multi-sensor data, this may result in colours that are not at all intuitive to the interpreter. Considering, *eg*, Figure 5.16(c), evidently it is very difficult to appreciate the characteristics of the original images from the various additive colour mixtures in such a colour composite. The algebraic manipulations of step 3 aim to overcome this problem. The transformations of step 3 have in common that they map the DN variations in the input images in one way or another on the perceptual attributes of human colour vision. These colour attributes are usually approached by using the IHS colour space presented in Section 5.2.1, which explains the frequent use of the RGB-IHS colour space forward and inverse transformations in image fusion.

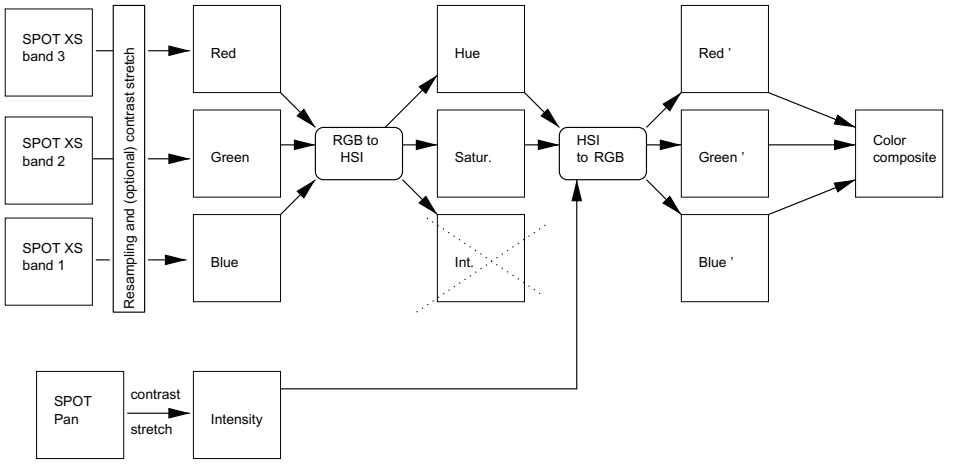


Figure 5.14: Procedure to merge SPOT panchromatic and multispectral data using RGB to IHS conversion and *vice versa*.

Most image processing systems, therefore, provide RGB-IHS colour space forward and inverse transformations as standard spectral enhancement tools.

Regardless of the colour space transformation selected, there are various ways in which a data set can be mapped to the perceptual attributes of the IHS space. Here we discuss only the most commonly used methods. Complementary information extraction by image fusion often exploits the so called 'image sharpening' principle, in which a single-band image of higher spatial resolution is merged with a multispectral band triplet of lower spatial resolution, in effect leading to a sharpened colour composite. This image fusion method is also named 'intensity substitution' or 'pan-sharpening' and is schematically illustrated in Figure 5.14. First a multispectral band triplet is transformed from

RGB to IHS space. Second, the intensity is replaced by the high spatial resolution image enhanced to a similar dynamic range by a linear contrast stretch or a histogram equalization. Third, the original hue and saturation and new intensity images are transformed back to RGB display for visualization on the computer monitor. There are many variants to this technique, including the contrast stretch of the saturation image or the mapping of two images on hue and intensity while setting saturation to a constant value. Usually, however, the hue image is left untouched, because its manipulation will result in a distorted representation of multispectral information. A simple algebraic operation to carry out an intensity substitution without saturation enhancement is known as the *Brovey transform*, written as:

$$R' = \frac{R}{I} I' \qquad G' = \frac{G}{I} I' \qquad B' = \frac{B}{I} I' \tag{5.3}$$

with

$$I = \frac{1}{\sqrt{3}}(R + G + B),$$

where R, G, B are the contrast stretched bands of the multispectral band triplet, I is the intensity of the multispectral band triplet, I' is the image substituted for the intensity and R', G', B' the band triplet of the fused data set.

An alternative arithmetic algorithm used for 'image sharpening' adds the high spatial resolution image in equal amounts to each spectral band in a technique

known as pixel addition ([4]):

$$R' = aR + bI' \qquad G' = aG + bI' \qquad B' = aB + bI' \tag{5.4}$$

with

$$a, b > 0 \qquad and \qquad a + b = 1,$$

where a and b are scaling factors to balance intensity sharpening from the high spatial resolution image versus the hue and saturation from the multispectral band triplet.

Next to colour space transformations and arithmetic combinations, statistical transforms such as principal component and regression analysis have been frequently used in image fusion, since they have the theoretical advantage of being able to combine a large number of images ([9]). In practice, however, fused images derived from many images are difficult to interpret.

The fourth and final step is essentially equivalent to the display of colour composites. Regardless of which fusion method is used, one needs to assure that the results are visualized as unambiguously as possible using the RGB space properties of the computer monitor. In practice this means that the perceptual colour attributes – intensity, hue and saturation – should be addressed proportionally to the dynamic ranges of the input images. Therefore, standard contrast enhancement procedures based on the histogram of each image may yield poorly optimized or even misleading displays. Because the DNs of each fused image

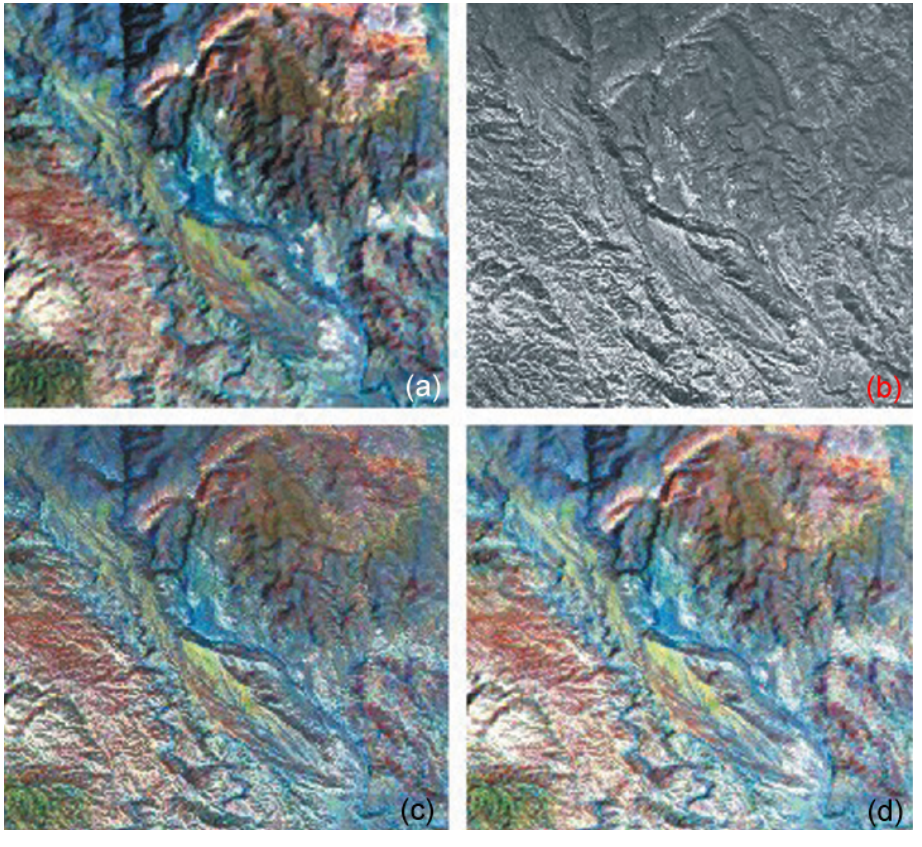


Figure 5.15: Fused images generated from bands 7, 3 and 1 of a Landsat-7 ETM subscene and an orthophoto mosaic of the Tabernas area, Southeast Spain. (a) colour composite of red = band 7, green = band 3 and blue = band 1; (b) orthophoto mosaic resampled to 5 metre GSD; (c) fused image of ETM bands, its intensity substituted with the orthophoto mosaic using RGB-IHS colour space transformations; (d) as (c) but with preservation of intensity of ETM bands by adding back 50 % of the original intensity to the intensity substitute, leading to better preservation of the spectral properties.

are actually composite values derived from two or more images, the contrast enhancement should be uniformly applied to each of the RGB images. This is to avoid ranges in hue out of proportion to the hue variations in a colour com-

posite image obtained from the multi-band images ([25]). Ideally, as shown in Figures 5.15 and 5.16, the colour composite image generated from the fused RGB channels should have the same hue range as the colour composite image generated from the original multispectral bands.

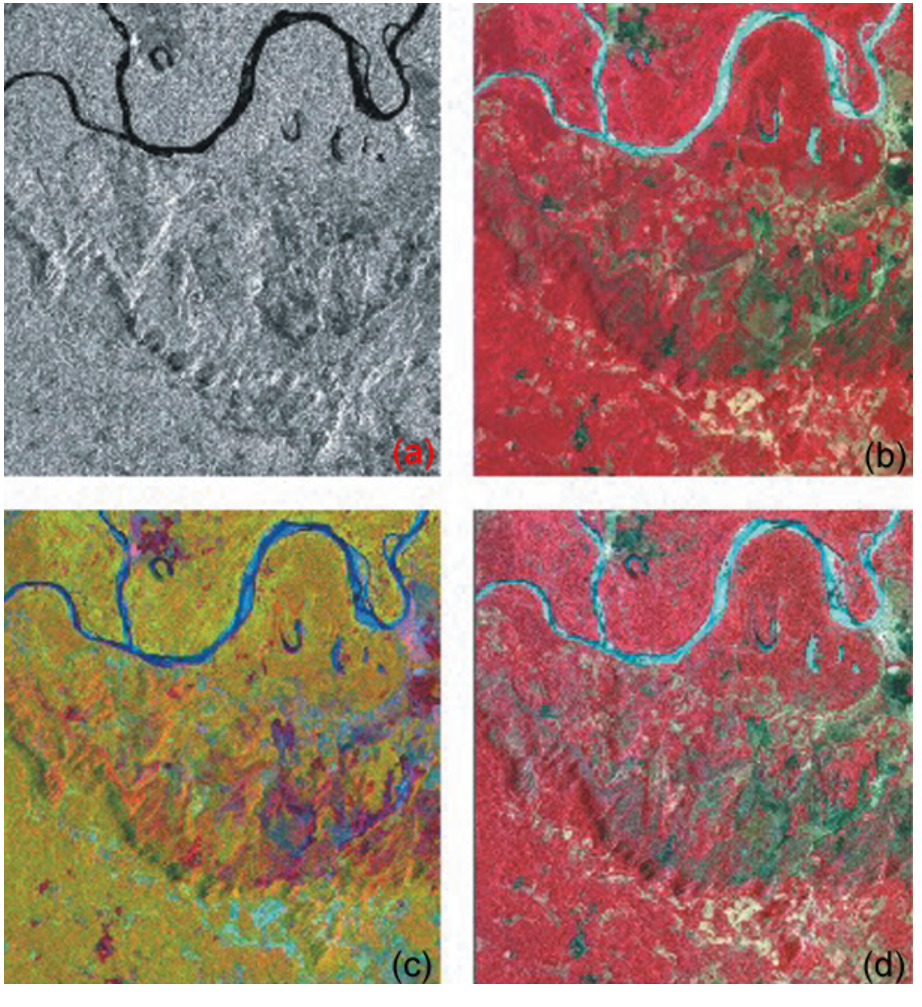


Figure 5.16: Fused images generated from ERS-1 SAR image and bands 3, 2 and 1 of a SPOT-2 subscene acquired over Amazon rain forest in Southeastern Colombia. (a) ERS1 image, (b) colour composite image of SPOT data, band 3 = red, band 2 = green and band 1 = blue; (c) colour composite image of ERS1 image = red, band 3 = green and band 2 = blue; (d) fused image of the three SPOT bands and the ERS1 image by using the pixel addition technique. Note that in the false colour composite the spectral information has been approximately preserved by using this technique.

Image fusion examples

In this section some examples are presented to illustrate various applications of image fusion and image processing details by which image fusion results can be optimized.

Figure 5.15 shows an example of an image sharpening application. Bands 7, 3 and 1 of a Landsat-7 ETM+ image acquired over an arid desert-like area in Southeastern Spain displayed as a colour composite in Figure 5.15(a), are co-registered and fused with an orthophoto mosaic (b) at a spatial resolution of 5 metres, by using the RGB-IHS colour space transformations. Figure 5.15(c) shows the image fusion result by substituting the intensity computed from the three contrast-enhanced ETM bands with the orthophoto mosaic. The high spatial resolution details of the orthophoto mosaic helps the interpreter to associate drainage patterns and terrain morphology with recent alluvial sediments and landforms in areas of badland erosion underlain by sedimentary rock units with contrasting spectral properties. Note that the hue and saturation of the colour composite in (a) have been preserved. The intensity information (*eg*, areas of high correlation among the ETM bands) however, has been lost in this image. This can be overcome by replacing the intensity by a weighted average of the intensity and the orthophoto mosaic, as illustrated in (d).

Figure 5.16 shows an image fusion application based on an ERS-1 synthetic aperture radar image (a) and SPOT-2 multispectral data (b) of an area covering the Amazon forest in southern Colombia. Figure 5.16(c) shows a colour composite image derived from these input data, where the ERS-1 image is displayed

through the red channel, SPOT band 3 through the green channel and SPOT band 2 through the blue channel. Note that, although this image shows an acceptable colour enhancement, the image colours cannot be intuitively related to the dynamic ranges of the input images, thereby hampering interpretation of the multispectral and backscatter properties of the SPOT and ERS-1 data, respectively. By contrast, Figure 5.16(d) shows an image fusion result which respects the perceptual attributes of IHS space. This image has been generated using the pixel addition technique explained above. Terrain morphology of the plateau in the west and the river flood plain, apparent on the ERS1 image are enhanced without loosing the false colour spectral information modulated by the three SPOT bands vital to the interpretation of vegetation cover.

Remember that in areas of high relief, distortions caused by layover and foreshortening (see Section 10.2.5) can be severe, leading to potential problems in pixel-to-pixel registration unless properly rectified.

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 computer monitors is the RGB space. When displaying an image on a screen (or as 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, and the use of image enhancement techniques.

The histogram, and the derived cumulative histogram, are the basis for contrast enhancement. Simple contrast enhancement methods are linear stretch and histogram equalization. Histogram operations can be specified by the transfer function. Smoothing or sharpening an image can be achieved by filtering. A linear filter on raster data can be specified by a kernel. Image enhancement by contrast stretching or filtering changes the radiometric properties of RS data; it is applied to obtain images better suited for visual interpretation.

Another type of radiometric correction (which is usually applied by RS data providers) is image restoration. Image restoration and atmospheric corrections, which convert DNs to radiances, are further detailed in Chapter 11. A simple radiometric correction, which can be performed as a linear contrast stretch operation, is haze correction. Image fusion can be seen as an extension of image enhancement with the purpose to exploit complementary characteristics of "multi-sensor" images.

Additional reading

[15], [3].

Questions

The following questions can help you to study Chapter 5.

1. How many possibilities are there to visualize a four-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 images?
4. Can you think of a situation in your own context where you would probably use filters to achieve easier interpretation of images?
5. Why is the effect of haze more pronounced in shorter wavelength bands?

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 the histogram of a digital image?

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.

6. Draw the transfer function for a haze correction by contrast stretching for the green channel of SPOT-5, assuming a skyradiance corresponding to DN=25.

Chapter 6

Geometric operations

6.1 Introduction

If you had not known before, by reading Chapter 3 you learned that the Earth has a spherical shape and that several clever men had devised transformations to map the curved Earth's surface to a plane. Through a map projection we obtain an image of the Earth's surface that has convenient geometric properties. We can, for instance, measure angles on a map and use these for navigation in the real world, or for setting out a designed physical infrastructure. Or if we use instead of a conformal projection like UTM an equivalent projection, we can determine the size of a parcel using the map - irrespective of where on the map the parcel is and at which elevation it is on the Earth. A remote sensor images the Earth's surface without knowledge of map projections, so we must not expect that remote sensing images have the same geometric properties as a map. Moreover, wherever a remote sensor takes an image, it merely records DNs. The DNs do not come with a label that would tell us where exactly on the Earth is the corresponding ground resolution cell (see Figure 6.1).

Luckily we get the DNs delivered in an orderly way, well arranged in rows and columns. The position of a point in the image is uniquely defined by the row and column number of the pixel that represents the point. Relating a pixel position in the image to the corresponding position on the ground is the purpose of 'georeferencing the image'. Once we have figured out what the geometric relationship is between points on the ground and the corresponding point in the RS image, we can transform any pixel position to a position on the ground. 'Position on the ground' can be defined either in a 3D terrain coordinate system or through a map projection in a 2D map coordinate system. By georeferencing an

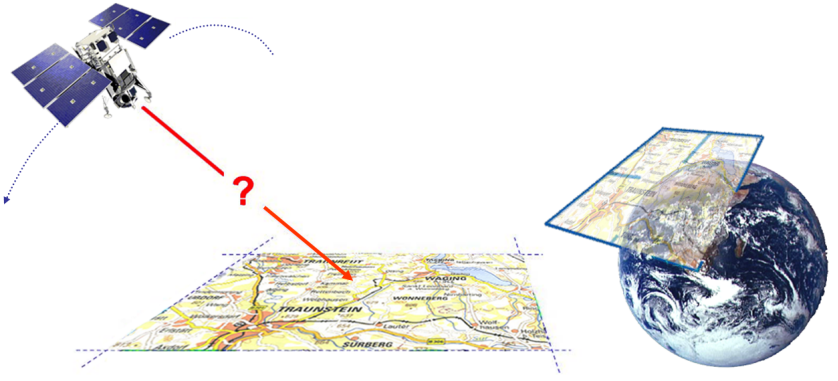


Figure 6.1: The problem of georeferencing a RS image.

image we solve two problems at the same time: (1) we can get map coordinates of features which we identify in the image, and (2) we implicitly correct for geometric distortions of the RS image if we compute correct map coordinates for any pixel. It takes georeferencing to turn raw RS data into geospatial data. After georeferencing (or speaking more generally: sensor orientation), we can:

- Measure in images to obtain 2D and 3D object descriptions. Mapping the place where we live has been our concern since thousands of years. Navigation and military purposes were the main triggers for topographic mapping and RS - photogrammetry more specifically - has made mapping much more efficient. Mapping is still a prime application of RS although qualitative earth observation is catching up quickly because of the exponential damage we are causing to our environment. We are interested in mapping our environment at different spatial and thematic resolutions and accuracies. For many applications 2D representations of objects (by points, lines, and areas) suffice. As long as certain conditions are met, we can ob-

tain such representations from a single image and simple georeferencing, which directly relates the RS image to the digital map. We need stereo images or multiple images for applications requiring 3D coordinates or better 2D coordinates for mapping scenes of large elevation differences or objects with large height differences. Sensor orientation must then be done by more rigorous approaches than 2D georeferencing.

- Combine an image with vector (digital map) data. Assume you would like to see how land property units relate to land cover. If you had a digital cadastral map, you could readily overlay the parcel boundaries on a RS image, *eg*, a scanned photograph, which nicely shows land cover.

- Compare or fuse different images and data sets in a GIS. You may wish to detect land cover changes with time, or merge a radar image with an image from an optical sensor to replace imaged clouds, or produce an image map with classified and accordingly symbolized roads. Relating or merging different RS images and/or map data can conveniently be done if all the different data sets do not differ in geometry, if they are all transformed to the same geometric reference system. Computationally producing a new image from a RS image such that it fits a specific map projection is often referred to as *geocoding*.

Georeferencing and geocoding are elaborated in Section 6.3. Section 6.4 introduces camera orientation for dealing with 3D information and gives an overview of mapping techniques that take into account the 3^{rd} dimension. Section 6.2 reviews elementary image distortions. Scanners yield images of complex geometry, which becomes more apparent the larger the FOV. Detailing scanner geometry is beyond the scope of this book; if you wanted to learn more about it, you

could consult [15]. The geometry of radar is also peculiar and not treated in Section 6.2 but in Chapter 10.

6.2 Elementary image distortions

Each sensor-platform combination is likely to have its own type of geometric image distortion. Here we only expose three very common types: (1) the effect of oblique viewing, (2) the effect of earth rotation, and (3) 'relief displacement'. Tilting a camera (see Figures 9.1 and 9.2) leads to images of non-uniform scale (Figure 6.2). Objects in the foreground appear larger than those farther away from the nadir. Earth rotation affects spaceborne scanner and line camera images that have a large spatial coverage. The resulting geometric image distortion (Figure 6.2) can easily be corrected by 2D georeferencing. Relief displacement shows up specifically in camera images of large scale if there is significant terrain relief or if there are high protruding objects.

Oblique view
Earth rotation

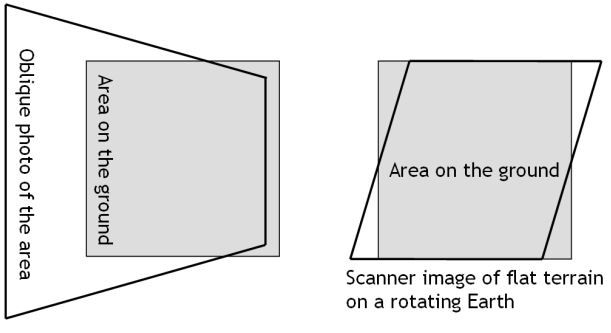


Figure 6.2: Examples of geometric image distortion.

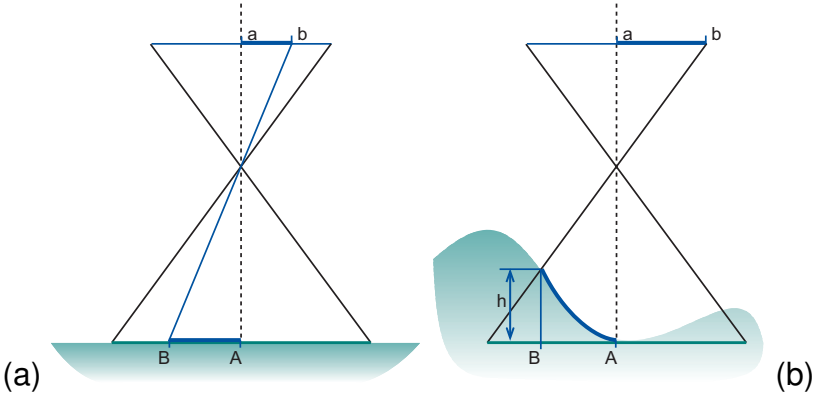


Figure 6.3: Illustration of the effect of terrain relief on the relationship between A–B on the ground and a–b in the image: (a) flat terrain, (b) significant elevation difference.

6.2.1 Relief displacement

A characteristic of most sensor systems is the distortion of the geometric relationship between the image and a conventional map of the terrain caused by elevation differences. This effect is most apparent in aerial photographs but also in images from spaceborne line cameras. The effect of relief displacement is illustrated in Figure 6.3 for a line camera. Consider the situation on the left, in which a true vertical image 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 size, respectively. In the left hand situation, by using the scale factor, we can compute $(A - B)$ from a measurement of $(a - b)$ in the image. In the right hand situation, there is significant terrain elevation difference. As you can now observe, the distance between a and b in the image 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 a central projection image but becomes increasingly prominent towards the edges of a camera image. This effect is called *relief displacement*: terrain points whose elevation is above or below the reference elevation are displaced in the image away from or towards the nadir point. The nadir, A, is the point on the ground directly beneath the sensor. The magnitude of displacement, δr (in mm), is approximated by:

Shifts due to height and elevation

$$\delta r = \frac{r \cdot h}{H}. \tag{6.1}$$

In this equation, r is the radial distance (in mm) from the nadir point, h (in m) is the terrain elevation above the reference plane, and H (in m) is the flying height

above the reference plane. The equation shows that the amount of relief displacement is zero at the nadir point, a, ($r = 0$) and largest at the edges of a line camera image and the corners of a frame camera image. Relief displacement is inversely proportional to the flying height.

Figure 6.4: Fragment of a large scale aerial photograph of the centre of Enschede. Note the effect of height displacement on the higher buildings.

In the case of scene of only barren land, we cannot see relief displacement. However, we can see relief displacement if there are protruding objects in the scene (it is then occasionally referred to as height displacement). Buildings and trees appear to lean outward, away from the nadir point on large scale photographs and images from very high resolution spaceborne sensors (Figure 6.4).

The main effect of relief displacement is that inaccurate or wrong map coordinates will be obtained when, *eg*, digitizing images without further correction. Whether relief displacement should be considered in the geometric processing of RS data depends on its impact on the accuracy of the geometric information derived from the images. Relief displacement can be corrected for if information about terrain relief is available (in the form of a DTM). The procedure is explained later in Section 6.4. It is also good to remember that it is relief displacement allowing us to perceive depth when looking at a stereograph and to extract 3D information from such images.

Effect of relief displacement

6.3 Two-dimensional approaches

In this section we consider the geometric processing of RS images in situations where relief displacement can be neglected. An example of such images is a scanned aerial photograph of flat terrain. For practical purposes, "flat" may be considered as $h/H < 1/1000$, though this also depends on project accuracy requirements; h stands for relative terrain elevation, H for flying height. For satellite RS images of medium spatial resolution, relief displacement usually is less than a few pixels in extent and thus less important, as long as near vertical images are acquired. The objective of 2D georeferencing is to relate the image coordinate system to a specific map coordinate system (Figure 6.5).

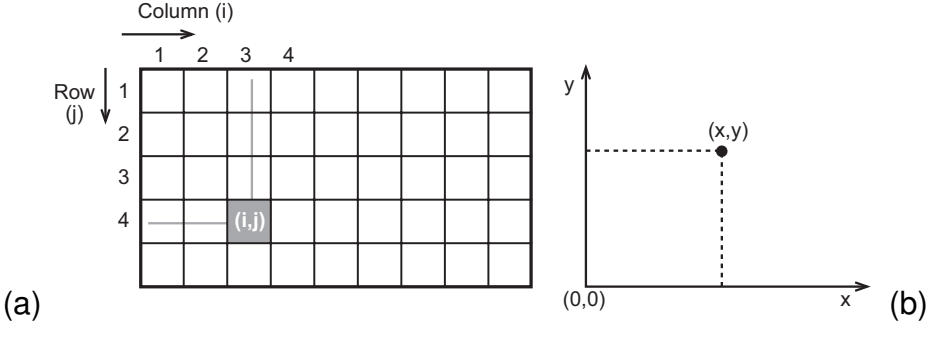


Figure 6.5: Coordinate system of (a) the image defined by its raster rows and columns, and (b) the map with x- and y-axes.

6.3.1 Georeferencing

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

Transformation

Using the above transformation, *eg*, the image position $(i = 3, j = 4)$ corresponds to map coordinates $(x = 18, y = 8)$. Once the transformation parameters have been determined, the map coordinates for every pixel can be calculated. This implies that we can superimpose on the image vector data that are given in the map coordinate system, or that we can store features by map coordinates when applying on-screen digitizing. Note that the image in the case of georeferencing remains stored in the original (i, j) raster structure and that its geometry is not altered. As we will see in Section 6.3.2, transformations can also be used to change the actual shape of an image and thus make it geometrically equivalent to the map.

The process of georeferencing involves two steps: (1) selection of the appropriate type of transformation and (2) determination of the transformation parameters. The type of transformation depends mainly on the sensor-platform system used. For aerial photographs (of flat terrain), the so-called 'projective transformation' models well the effect of pitch and roll (see Figures 9.1, 6.2 and 6.8). A more general type 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

Type of transformation

(i, j) as follows:

$$x = a + bi + cj \tag{6.2}$$

$$y = d + ei + fj \tag{6.3}$$

Equations 6.2 and 6.3 require that six parameters (a to f) be determined. The transformation parameters can be determined by means of *ground control points* (GCPs). GCPs are points that can be clearly identified in the image and in the target map. The target map could be a topographic map or another image that has been transformed to the wanted map projection system before. The operator then needs to identify corresponding points on both images. It depends on the image and map scale what suitable points are. Typical examples are: road crossings, crossings of waterways, salient morphological structures, etc. Another possibility is to identify points in the image and to measure the coordinates of these points in the field, *eg*, by GPS and then transform those to map coordinates. It is important to note that it can be quite difficult to identify good GCPs in an image, especially in lower-resolution RS images. Once a sufficient number of GCPs has been measured in the image and on the map (or in the field), the software sets up the two equations for each point and solves the system of equations. The outcome are the values for the parameters a to f of the Equations 6.2 and 6.3 and quality indications.

Ground control points

To solve the above equations, only three GCPs are required; however, you should use more points than the strict minimum. Using merely the minimum number of points for solving the system of equations would obviously lead to a wrong

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 6.1: 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.

Number of GCPs

transformation if you made an error in one of the measurements. Including more points for calculating the transformation parameters enables the software to also compute the error of the transformation. Table 6.1 provides an example of the input and output of a georeferencing computation in which 5 GCPs have been used. Each GCP is listed with its image coordinates (i, j) and its map coordinates (x, y) .

The software performs a 'least-squares adjustment' to determine the transformation parameters. The least squares adjustment ensures an overall best fit of image and map. We use the computed parameter values to calculate coordinates (x_c, y_c) for any image point (pixel) of interest:

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

and

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

For example, for the pixel corresponding to GCP 1 (i=254 and j=68), we can calculate transformed image coordinates x_c and y_c as 958.552 and 154.935, respectively. These values deviate slightly from the input map coordinates (as measured on the map). The discrepancies between measured and transformed coordinates of the GCPs are called residual errors (*residuals* for short). The residuals are listed in the table as d_x and d_y. Their magnitude is an indicator of the quality of the transformation. The residual errors can be used to analyze whether all GCPs have correctly been identified and/or been measured accurately enough.

Residuals

The overall accuracy of a transformation is either stated in the accuracy report of the software in terms of variances or as *Root Mean Square Error* (RMSE). The RMSE in x-direction, m_x, is calculated from the residuals (at check points) using the following equation:

Accuracy of georeferencing

$$m_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \delta {x_i}^2}. \tag{6.4}$$

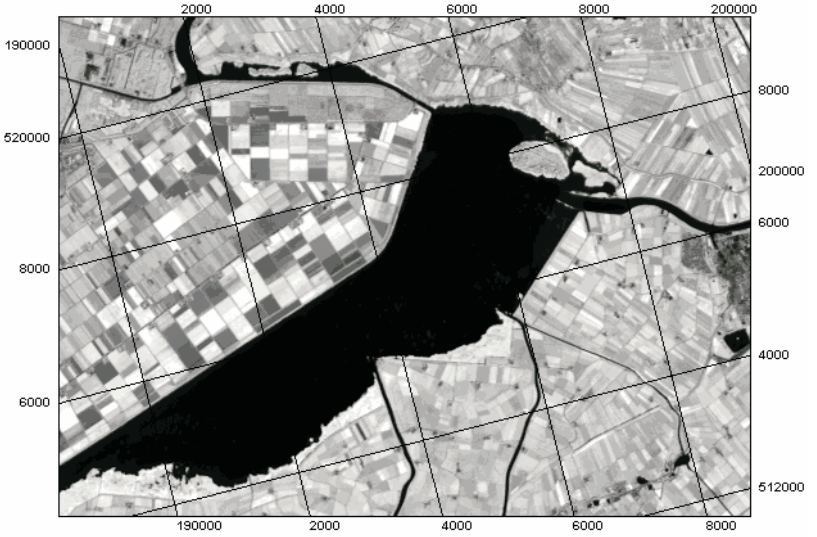
For the y-direction, a similar equation can be used to calculate m_y. The overall RMSE, m_p, is calculated by:

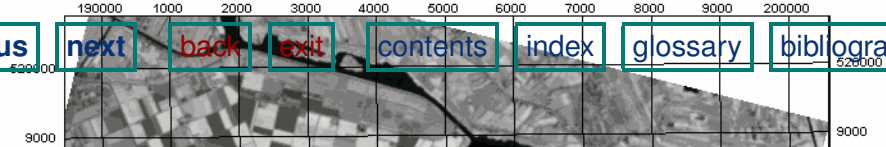
RMSE

$$m_p = \sqrt{{m_x}^2 + {m_y}^2}. \tag{6.5}$$

For the example data set given in Table 6.1, the residuals have been calculated: the respective m_x, m_y and m_p are 1.159, 0.752 and 1.381. The RMSE is a convenient overall accuracy measure, but it does not tell us which parts of the image

are accurately transformed and which ones are not. Note also that the RMSE is only valid for the area that is bounded by the GCPs. In the selection of GCPs, therefore, points should be well distributed and include locations near the edges of the image.


2000
4000
6000
8000
200000
190000
520000
8000
6000
8000
8000
200000
6000
4000
512000
190000
2000
4000
6000
8000


190000
1000
2000
3000
4000
5000
6000
7000
8000
9000
200000
520000
526000
9000
9000

6.3.2 Geocoding

The previous section explained that two-dimensional coordinate systems, *eg*, an image system and a map system, can be related using a geometric transformation. This georeferencing approach is useful in many situations. However, in other situations a 'geocoding' approach, in which the row-column structure of the image is also transformed, is required. Geocoding is required when different images need to be combined or when the images 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 illustrated by Figure 6.6. Distinguishing georeferencing and geocoding is conceptionally useful, but to the casual software user it is often not apparent whether only georeferencing is applied in certain image manipulations or also geocoding.

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: (1) each new raster element is projected (using the transformation parameters) onto the original image and (2) a DN for the new pixel is determined and stored.

We can consider the relationship between transformation and resampling in Figure 6.7. When transforming the original raster, its shape changes. Since raster data are stored in a regular row–column pattern, we need to calculate DNs for the pixel pattern of the new, corrected image. We say we have to *resample* the original image. Imagine that the transformed image is overlaid with a grid corresponding to the output image and the pixel values for the new image are de-

Resampling

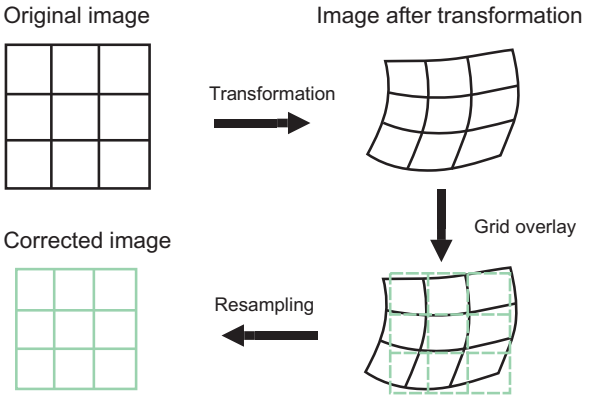


Figure 6.7: Illustration of the transformation and resampling process. Note that the 'image after transformation' is a conceptual illustration, since the actual pixel shape cannot change.

termined. The shape of the image to be resampled depends on the type of transformation used.

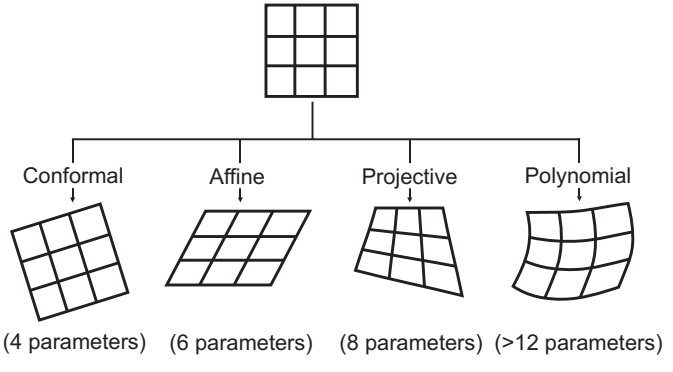


Figure 6.8: Illustration of different image transformation types, and the number of required parameters.

Figure 6.8 shows four transformation types that are frequently used in RS. The

types shown increase in complexity and parameter requirement, from left to right. In a conformal transformation the image shape, including the right angles, are retained. Therefore, only four parameters are needed to describe a shift along the x and the y axis, a scale change, and the rotation. However, if you want to geocode an image to make it fit with another image or map, a higher-order transformation may be required, such as a projective transformation or polynomial transformation.

Conformal transformation

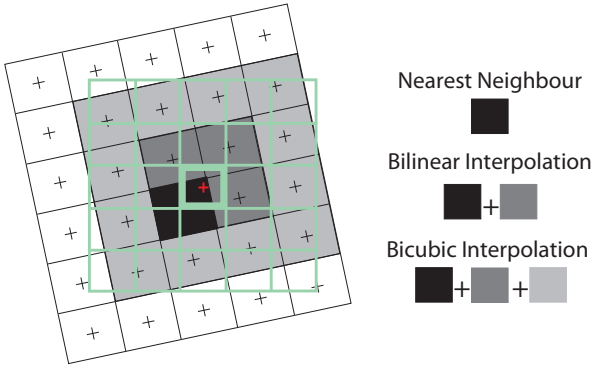


Figure 6.9: Principle of resampling using nearest neighbour, bilinear, and bicubic interpolation.

Regardless of the shape of the transformed image, a resampling procedure as illustrated by Figure 6.9 is used. As the orientation and size of the input raster and the required output raster are different, there is not a one-to-one relationship between the cells of these rasters. Therefore, we have to apply interpolation to compute a pixel value for a ground resolution cell where it has not been measured. For resampling we only use very simple interpolation methods. The main methods are: nearest neighbour, bilinear interpolation, and bicubic interpola-

Interpolation

tion (Figure 6.9). Consider the green grid to be the output image to be created. To determine the value of the centre pixel (bold), in *nearest neighbour* the value of the nearest original pixel is assigned, the value of the black pixel in this example. Note that always the respective pixel centres, marked by small crosses, are used for this process. When using *bilinear interpolation*, a weighted average is calculated from the four nearest pixels in the original image (dark grey and black pixels). In *bicubic interpolation*, a weighted average of the values of 16 surrounding pixels (the black and all grey pixel) is calculated [30]. Note that some software uses the terms "bilinear convolution" and "cubic convolution" instead of the terms introduced above.

Choice of method

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. Nearest neighbour resampling can lead to the edges of features to be offset in a step-like pattern. However, since the value of the original cell is assigned to a new cell without being changed, all spectral information is retained, which means that the resampled image is still useful in applications such as digital image classification (see Chapter 8). The spatial information, on the other hand, may be altered in this process, since some original pixels may be omitted from the output image, or appear twice. Bilinear and bicubic interpolation reduce this effect and lead to smoother images; however, because the values from a number of cells are averaged, the radiometric information is changed (Figure 6.10).

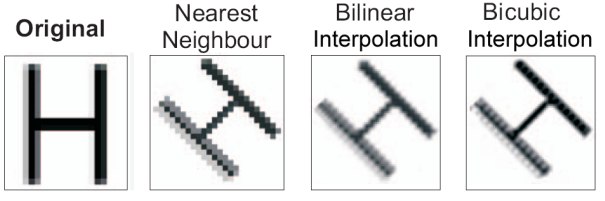


Figure 6.10: The effect of nearest neighbour, and bilinear and bicubic resampling on the original data.

6.4 Three-dimensional approaches

In mapping terrain, we have to consider its vertical extent (elevation and height) it two types of situation:

- We want 2D geospatial data, describing the horizontal position of terrain features, but the terrain under considerations has large elevation differences. Elevation differences in the scene cause relief displacement in the image. Digitizing image features without taking into account relief displacement causes errors in computed map coordinates. If the positional errors were larger than tolerated by the application (or map specifications), we should not go for simple georeferencing and geocoding.
- We want 3D data.

When bothering about mapping terrain with an increasing degree of refinement we have to first clarify what we mean with 'terrain'. *Terrain* as described by a topographic map has two very different aspects: (1) there are agricultural fields and roads, forests and waterways, buildings and swamps, barren land and lakes, and (2) there is elevation changing with position - at least in most regions on Earth. We refer to land cover, topographic objects, etc as *terrain features*; we show them on a (2D) map as areas, lines, and point symbols. We refer to the shape of the ground surface as *terrain relief*; we show it on a topographic map by contour lines and/or relief shades. A *contour line* on a topographic map is a line of constant elevation. Given contour lines, we can determine the elevation at any arbitrary point by interpolating between contour lines. We could digitize

Terrain relief

the contour lines of a topographic map. The outcome would be a data set consisting of (X, Y, Z) coordinates of many points. Such a data set is the basis of a 'digital terrain relief model' (DTM). We then need a computer program to utilize such a coordinate list of points of the ground surface to compute elevation at any horizontal position of interest and derive other terrain relief information as, *eg*, slope and aspect. The idea of a DTM was developed in the late 1950'ies at the MIT when aiming at computerizing highway design. 50 years later we use DTMs in all geosciences with a wide range of applications and we have remote sensors specifically built to supply us with data for DTMs (SRTM, SPOT-5 HRS, etc). One of the applications of a DTM is to accomplish 'digital monoplotting' and 'orthoimage' production as outlined later in this section.

Terrain models

A *DTM* is a digital representation of terrain relief, a model of the shape of the ground surface. We have a variety of sensors at our disposal that can provide us with 3D data: line cameras and frame cameras, laser scanners, and microwave radar instruments. They all can produce (X, Y, Z) coordinates of terrain points, but not all of the terrain points will be points on the ground surface. Consider a stereo pair of photographs, or a stereo pair from SPOT-5 of a tropical rain forest. Will you be able to see the ground? Obtained coordinates will pertain to points on the ground only in open terrain. Since a model based on such data is not a DTM, we refer to it as digital surface model (DSM). The difference between DTM and DSM is illustrated by Figure 6.11; we need to filter DSM data to obtain DTM data.

In terrain modelling, it is handy to choose the coordinate system such that Z stands for elevation. If we model a surface digitally by nothing else than eleva-

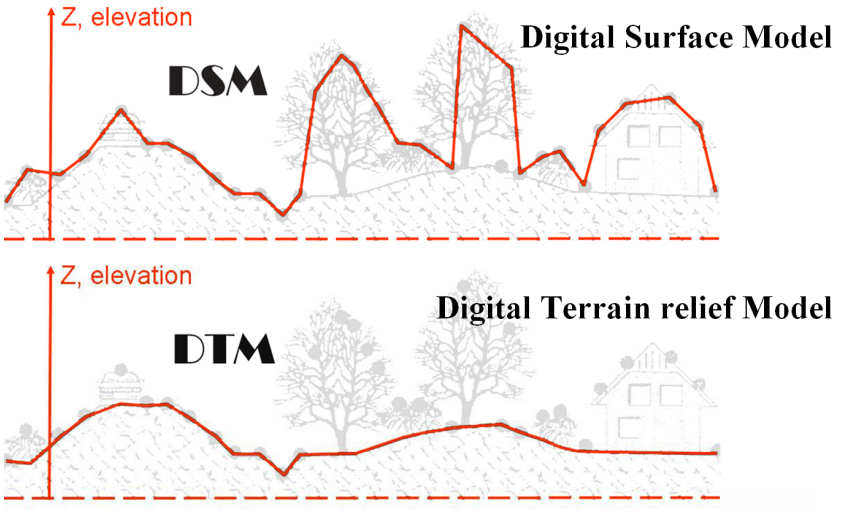


Figure 6.11: The difference between DTM and DSM.

tion values Z at horizontal positions (X, Y), why not call such a model a *digital elevation model* (DEM). The term DEM was introduced in the 1970'ies with the purpose to distinguish the simplest form of terrain relief modelling from more complex types of digital surface representation. The term DEM was original exclusively used for raster representations (elevation values given at the intersection nodes of a regular grid). Note that both a DTM and DSM can be a DEM and, moreover, 'elevation' would not have to relate to terrain but could relate to some subsurface layer such as ground water level, or some soil layer, or the ocean floor. Unfortunately you will find in literature perturbed uses of the terms as introduced above, in particular, DEM is often used carelessly. Worth mentioning in this context is also the misuse of "topography" as synonym for terrain relief.

6.4.1 Orientation

The purpose of 'camera orientation' is to obtain the parameter values for transforming terrain coordinates (X, Y, Z) to image coordinates and vice versa. In solving the orientation problem, we assume that any terrain point and its image lie on a straight line, which passes through the projection centre (*ie*, the lens). This assumption about the imaging geometry of a camera is called *collinearity* relationship and does not consider that atmospheric refraction has the effect of slightly curving imaging rays. We solve the problem of orienting a single image with respect to the terrain in two steps: 'interior orientation' and 'exterior orientation'.

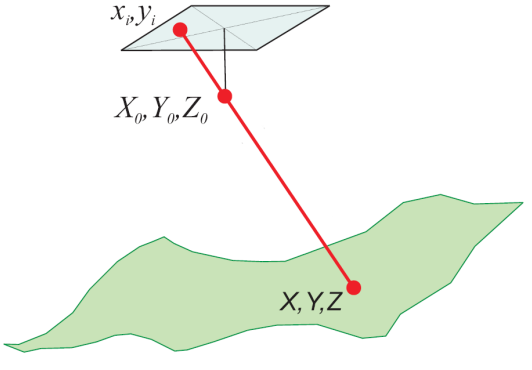


Figure 6.12: Illustration of the collinearity concept, where image point, lens centre and terrain point all lie on one line.

Orienting a single image as obtained by a line or frame camera

The *interior orientation* determines the position of the projection centre with respect to the image. The problem to solve is different for digital cameras and film cameras. In the case of a digital camera (line or frame camera, spaceborne or aerial camera) the position of the projection centre with respect to the CCD array does not change from image to image, unless there were extreme temperature or pressure changes. For standard applications we can assume that the position of the projection centre does not change when defined in the row-column system of the digital image. Two parameters are needed, the 'principal distance' and the 'principal point' (see Figure 6.13); they are determined by camera calibration. The *principal distance* is the geometrical equivalent of the focal length (which is physical property of a lens). The *principal point* is the intersection point of the perpendicular from the projection centre with the image plane. The camera calibration report states the row (and column) number of the principal point and the principal distance. When using digital photogrammetric software for working with images from digital aerial cameras, the user only has to enter the principal distance, the position of the principal point, and the pixel size to define the interior orientation. In the case of a film camera, the position of the principal point is only fixed in the camera and determined by camera calibration with respect to the fiducial marks. When scanning photographs (see Section 9.6), the principal point will have different positions in the row-column system of each image, because each image will be placed slightly differently in the scanner. Therefore, you have to measure for every image the imaged fiducial marks and calculate the transformation onto the fiducial marks as given by the calibration report. Digital photogrammetric software can then relate row-column coordinates of any image point to image/camera coordinates (x, y) (see Figure 6.13).

Interior orientation

Principal distance

Principal point

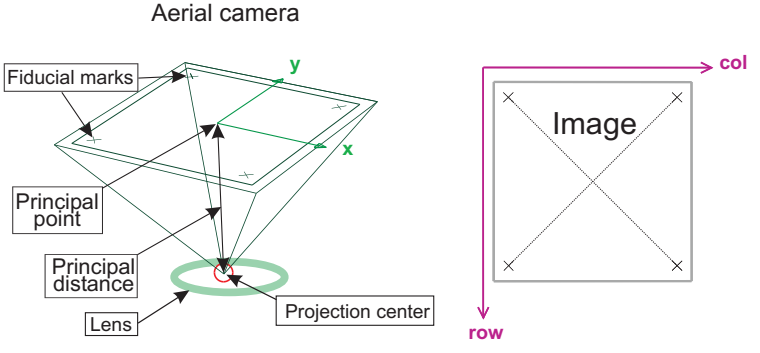


Figure 6.13: Inner geometry of a camera and the associated image.

The *exterior orientation* determines the position of the projection centre with respect to the terrain and also the attitude of the sensor. Frame cameras (film as well as digital ones) acquire the entire image at once, thus three coordinates for the projection centre (X, Y, Z) and three angles for the attitude of the sensor (angles relating to roll, pitch, and yaw) are sufficient to define the exterior orientation of an entire image. Line cameras yield images where each image line has its own projection centre and the sensor attitude may change from one line to the next. Satellites have a very smooth motion and airborne line cameras are mounted on a stabilized platform so that the attitude change is gradual and small. Mathematically we can model the variation of an attitude angle through a scene by a low degree polynomial. Images from modern high resolution line cameras on satellites come with 'rational polynomial coefficients' (RPC). The rational polynomials define by good approximation the relationship between image coordinates of an entire frame (in terms of row-column pixel positions) and terrain coordinates (usually in terms of Lat-Lon and ellipsoidal height of

Exterior orientation

Sensor position and attitude

RPC

WGS84). The nice thing is that the RPC are understood by RS software such as ERDAS thus taking care of interior and exterior orientation. For cases where RPC are not given or considered not accurate enough, you need GCPSs. For aerial surveys we can solve the exterior orientation in one of the following ways:

- *Indirect camera orientation*: identify GCPs in the image and measure the row-column coordinates; acquire (X, Y, Z) coordinates for these points, *eg*, by GPS or a sufficiently accurate topographic map; use adequate software to calculate the exterior orientation parameters, after having completed the interior orientation.
- *Direct camera orientation*: make use of GPS and IMU recordings during image acquisition employing digital photogrammetric software.
- *Integrated camera orientation*, which is a combination of (a) and (b).

For images from very high resolution sensors such as Ikonos or QuickBird, adding one GCP can already considerably improve the exterior orientation as defined by the RPC. For Cartosat images it is advisable to improve the exterior orientation by at least five GCPs. For orienting a frame camera image you need at least three GCPs (unless you also have GPS and IMU data). After orientation, you can use the terrain coordinates of any reference point and calculate its position in the image. The differences between measured and calculated image coordinates (the residuals) allow you to estimate the accuracy of orientation. As you may guess, advanced camera/sensor/image orientation is a topic for further studies.

Orienting a stereo-image pair obtained by a frame camera

The standard procedure is to individually orient each image. In the case of images stemming from a frame camera we can, however, readily make use of the fact that both images partly cover the same area. Instead of doing two independent exterior orientations we can better first do a 'relative orientation' of the two images, followed by an 'absolute orientation' of the pair to the terrain coordinate system. The *relative orientation* will cause the imaging rays of corresponding points to intersect more accurately than if we orient one image without the knowledge of the other image. We do not have to measure points with known terrain coordinates to solve the relative orientation problem; we only need to measure image coordinates of corresponding points in the two images - after the individual interior orientations have been established. Measuring of corresponding points can even be done automatically (by 'image matching', see below). For absolute orientation at least three GCPs are needed. The idea of splitting up the exterior orientation into a relative orientation and absolute orientation is also used for orienting a whole block of overlapping images, not only two. The advantage is that we only need few GCPs (which are usually expensive to acquire) and still obtain accurate transformation parameters for each image. The method is known as 'aerial triangulation'.

Relative orientation

Absolute orientation

6.4.2 Monoplotting

Suppose you need to derive accurate planimetric coordinates of features from a single aerial photograph expressed in a specific map projection. This can be achieved for flat terrain using a vertical photograph and a georeferencing approach. Recall from the earlier discussion of relief displacement (Figure 6.3) how elevation differences lead to distortions in the image, preventing the use of such data for direct measurements. Therefore, if there is significant terrain relief, the resulting relief displacement has to be corrected for. For this purpose the method of *monoplotting* has been developed, with a major research input of ITC.

3D data from a single image

Monoplotting is based on the reconstruction of the position of the camera at the moment of image exposure with respect to the GCPs, *ie*, the terrain. This is achieved by identification of several (at least four) GCPs, for which both the photo and map coordinates are known. Information about the terrain relief is supplied by a DTM of adequate accuracy. The DTM should be given in the required map projection system and the elevations should be expressed in an adequate vertical reference system. When digitizing features on the photograph, the computer uses the DTM to calculate the relief displacement for every point and corrects for it (Figure 6.14). A monoplotting approach is possible by using a hardcopy image on a digitizer tablet, or by on-screen digitizing on a computer monitor. In the latter situation, vector information can be superimposed on the image to update the changed features. Note that monoplotting is a (real time) correction procedure and does not yield a new image, *ie*, no resampling is carried out.

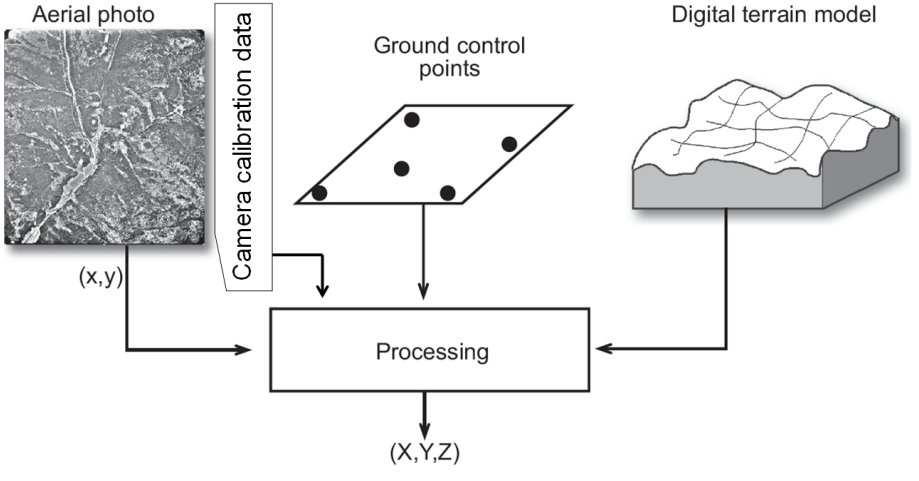


Figure 6.14: The process of digital monoplotting enables accurate determination of terrain coordinates from a single aerial photograph.

6.4.3 Orthoimage production

Geocoding an image of rough terrain

Monoplotting can be considered a georeferencing procedure that incorporates corrections for relief displacement, without involving any resampling. For some applications, however, it is useful to actually correct the photograph or RS image, taking into account the effect of terrain relief. In such cases the image should be transformed and resampled (making use of a DTM) into a product with the geometric properties of a specific map projection. Such an image is called an *orthophoto*, or more generally *orthoimage*.

The production of orthophotos is quite similar to the process of monoplotting. Consider a scanned aerial photograph. First, the photo is oriented using ground control points. The terrain elevation differences are modelled by a DTM. The computer then calculates the position in the original photo for each output pixel. Using one of the resampling algorithms, the output value is determined and stored in the required raster. The result is geometrically equivalent to a map, *ie*, direct distance or area measurements on the orthoimage can be carried out.

6.4.4 Stereo restitution

After relative orientation of a stereo pair, we can exploit the 3D impression gained from the stereo model to make measurements in 3D. The measurements made in a stereo model make use of the phenomenon called *parallax* (Figure 6.15). Parallax refers to the fact that an object photographed from different camera locations (*eg*, from a moving aircraft) has different relative positions in the two images. In other words, there is an apparent displacement of an object as it is observed from different locations. Figure 6.15 illustrates that points at two different elevations, regardless of whether it is the top and bottom of a hill or of a building, experience a relative shift. The measurement of the difference in position is a basic input for elevation calculations. We could use stereo restitution to measure (X, Y, Z) coordinates of many points, this way obtaining data for a DTM. This is both a boring and an error prone process. Digital photogrammetric software can do this job automatically to a high degree of success - after some 30 years of research and development - using 'image matching'. Manual measurements are then only needed as supplement for difficult areas. The main purpose of stereo restitution is to collect 3D vector data of objects. Different from monoplotting, elevation is directly measured and not interpolated from a DTM; hence the coordinate accuracy can be higher. Another main advantage of stereo restitution is the better image interpretability.

3D data from a stereo pair

Image matching

A stereo model enables parallax measurement using a special 3D cursor. If the stereo model is appropriately oriented, the parallax measurements yield (X, Y, Z) coordinates. 'Analogue' systems use hardcopy images and perform the computation by mechanical, optical or electrical means. 'Analytical' systems also

Analogue, analytical, digital systems

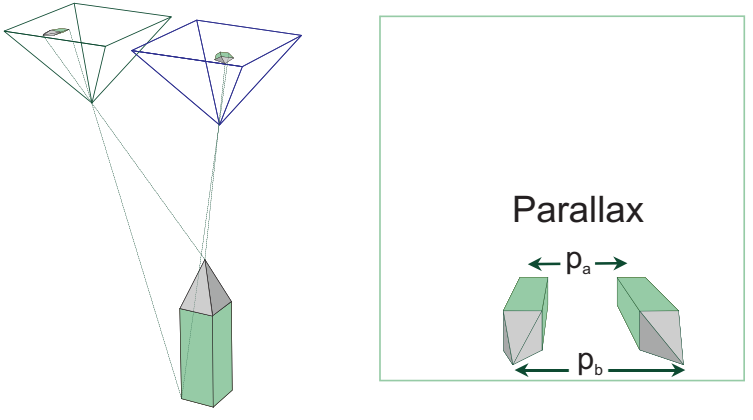


Figure 6.15: The same building is observed from two different positions. Because of the height of the building the positions of the building top and base relative to photo centres are different. This difference (parallax) can be used to calculate its height.

use hardcopy images, but do the computation digitally, while in modern 'digital' systems, both the images and the computation are digital. Using digital instruments we can not only make a few spot elevation measurements, but generate automatically a complete DSM for the overlapping part of the two images. Recall, however, that reliable elevation values can only be extracted if the orientation steps were carried out accurately, using reliable ground control points.

Summary

This chapter has introduced the basic geometric operations for making use of RS data. A basic concern in dealing with RS data is terrain relief, which can be neglected for some applications (2D approaches) or must taken into account (3D approaches). In both approaches there is a possibility to keep the original 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 makes use of parallax differences as observed in stereo images to measure the vertical dimension and obtain correct horizontal coordinates of terrain points.

Additional Reading

[15], [17].

Questions

The following questions can help you to study Chapter 6.

1. Think of two situations in which images are used and in which you need to take relief displacement into account.
2. For a transformation of a specific image to a specific map, an m_p error of two times the GSD is given. What additional information do you need to assess the accuracy of the transformation?
3. Calculate the map position (x, y) for image position $(10, 20)$ using the two following equations: $x = -10 + 5i - j$ and $y = 5 + 2i + 2j$

The following are typical exam questions:

1. Compare image and map coordinate systems (make a figure with comments).
2. What is the purpose of acquiring stereo-image pairs?
3. What are ground control points used for?
4. Explain the purpose of monoplotting. What inputs do you need?

Chapter 7

Visual image interpretation

7.1 Introduction

How to extract information from images? In general, information extraction methods from remote sensing images can be subdivided into two groups:

- Information extraction based on visual image interpretation. Typical examples of this approach are visual interpretation methods for land use or soil mapping. Also acquiring data for topographic mapping from aerial photographs is still largely 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, digital image classification and calculation of surface parameters. Digital image classification is introduced in Chapter 8.

The most intuitive way to extract information from remote sensing images is by visual image interpretation, which is based on our ability to relate colours and patterns in an image to real world features. Chapter 5 has explained different methods used to visualize remote sensing data. We can interpret images displayed on a computer monitor or images given in a hardcopy form. How to convey the findings of an interpreter to somebody else? In everyday life we do it verbally but since thousands of years we also do it by mapping. We used to overlay a transparency on a photograph and outline an area which we recognized to have the property of interest. Doing so for all features of interest in a scene we obtained a map. The digital variant of this approach is to digitize - either on-screen or using a digitizer tablet if we have a hardcopy image - points,

Mapping

lines, and areas and label these geometric entities to convey thematic attributes. This way we can obtain, *eg*, a map of all vineyards in a certain area and the roads and tracks leading to them. Instead of interpreting and digitizing a single image, we can use a stereo-image pair. The interpretation process is the same, we only need special devices for stereoscopic display and viewing, and equipment, which allows us to properly measure in a stereogram.

Visual image interpretation is not as easy as it may seem before having a closer look; it requires training. Yet our eye-brain system is very well capable of doing the job. That visual interpretation is in fact an extremely complex process, we discovered when we tried to let a computer do image interpretation. Research on 'image understanding' is still continuing and it has helped us to conceptionalize human vision and interpretation.

Section 7.2 explains the basics of how we accomplish recognizing features and objects in images. Visual image interpretation is used to produce geospatial data 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. Actual image interpretation is application specific, adhering, however, to a common approach. Section 7.3 describes the general, practical approach. Aspects of assessing the quality of the outcome of an interpretation are treated in Section 7.4.

7.2 Interpretation fundamentals

7.2.1 Human vision

In Chapter 5 our perception of colour was explained. *Human vision* goes a step beyond the perception of colour: it deals with the ability of a person to draw conclusions from visual observations. In analyzing an image, typically you are somewhere between the following two situations: direct and 'spontaneous recognition', or using several clues to draw conclusions by a reasoning process (*ie*, 'logical inference').

Spontaneous recognition refers to the ability of an interpreter to identify objects or features at first glance. Consider Figure 7.1. 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 what they see on an aerial photo to the terrain features of their place of living (because of 'scene knowledge'). The quote from people that are shown an aerial photograph for the first time "I see because I know" refers to spontaneous recognition.

Spontaneous recognition

Logical inference means that the interpreter applies reasoning. In the reasoning the interpreter uses acquired professional knowledge and experience. Logical inference is, for example, concluding that a rectangular shape is a swimming pool because of its location in a garden near to the house. Sometimes logical inference alone is insufficient to interpret images; then field observations are required (Section 7.4). Consider the aerial photograph in Figure 7.2. Are you 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 ele-

Logical inference

ments of a house or settlement.

Figure 7.1: RS image of Antequera area in Spain, the circular features are pivot irrigation systems. The area imaged is 5 km wide.

Figure 7.2: Mud huts of Labbezanga near the Niger river. Photo by Georg Gerster, 1972.

7.2.2 Interpretation elements

We need a set of terms to express characteristics of an image, which we use when interpreting the image. 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 seven interpretation elements are distinguished: tone/hue, texture, pattern, shape, size, height/elevation, and location/association.

- *Tone* is defined as the relative brightness in a black-and-white image. *Hue* refers to the colour as defined in the IHS colour space. Tonal variations are an important interpretation element. The tonal expression of objects in the image is directly related to the amount of light (or other forms of EM energy) reflected (or emitted) 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 imaged terrain and also to the bands selected for visualization (see Chapter 5). 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).
- *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, etc. Texture can often be related to ter-

rain surface 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.

- *Pattern* refers to the spatial arrangement of objects and implies the characteristic repetition of certain forms or relationships. Pattern can be described by terms such as concentric, radial, checkerboard, etc. Some land uses have specific and characteristic patterns when observed from the air or space. We 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.

- *Shape* or form characterizes many objects visible in the image. Both the two-dimensional projection of an object as shown on a map and the height of an object influence the shape of the imaged object. The shape of objects often helps to identify them (built-up areas, roads and railroads, agricultural fields, etc).

- *Size* of objects can be considered in a 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 the width determines the road type, *eg*, primary road, secondary road, etc.

- *Height* differences are important for distinguishing between different vegetation types, building types, etc. *Elevation* differences provide us with cues in geomorphological mapping. We need a stereogram and stereoscopic viewing to observe height and elevation. Stereoscopic viewing facilitates interpretation of both natural and man-made features.

- *Location/association* refers to the the situation in the terrain or in relation to its surrounding. A forest in the mountains is different from a forest close to the sea or near the river in the lowland. A large building at the end of a number of converging railroads is likely to be a railway station—we should not expect a hospital at such a location.

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 group of adjacent pixels, not for a single pixel. 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 digital image classification (Chapter 8), only hue is utilized, which explains the limitations of automated methods compared to visual image interpretation.

7.3 Mapping

7.3.1 Interpretation

The assumption in mapping with the help of remote sensing images is that areas that look homogeneous in the image will have similar features on the ground. The interpretation process consists of delineating areas that internally appear similar and at the same time different from other areas. Making an interpretation of only one aerial photograph or a small part of an image from a spaceborne sensor seems quite simple. You have the overview of the entire area at all times and can easily compare one unit to another and decide if they are the same or different. However working with many photographs and also with several people will require a good definition of the units to be delineated.

The definition of units is based on what can be observed in the image. Different interpretation units can be described according to the interpretation elements. After establishing what the features are on the ground, 'interpretation keys' can be constructed. Based on these keys an interpretation of features can be made. These features are again described in terms of interpretation elements. If knowledge of the area is lacking (not yet available), you could also start an interpretation only based on interpretation elements (Figure 7.3). After fieldwork it will become clear, what the units represent on the ground.

Interpretation key

Prior to the delineation of the units a legend is constructed based on interpretation elements. The legend can be presented in the form of a table, where each element type is represented by a column. In Table 7.1 a theoretical example of such a description of a legend is presented. In this legend the "unit number" represents a yet unknown feature type and the corresponding row elements will

Interpretation legend

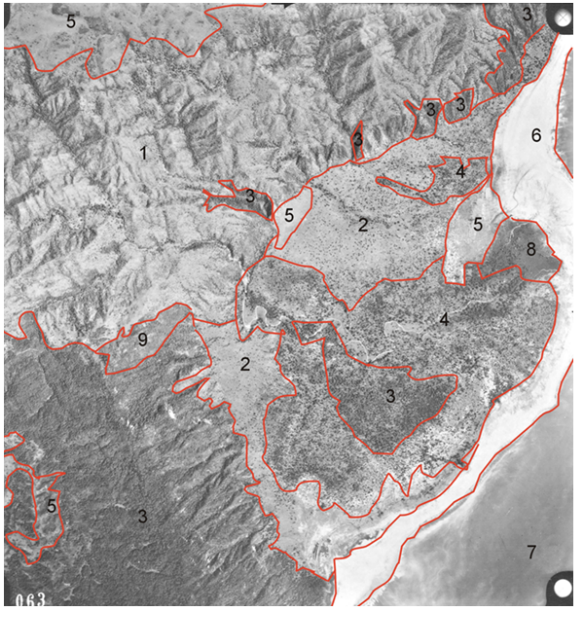


Figure 7.3: Example of an interpretation Manyara, Tanzania.

be used to identify that feature type.

While preparing the legend, consider that distinguishing units can be based on the difference in one element only or on differences of several elements. For example, consider Unit 1: its tone is black and all other units have a tone grey or white. In this case there is no need to define all the other elements for Unit 1. In the example of Figure 7.3 some units are different in texture. There are areas with smooth and rough texture. The rough texture areas are further differentiated according to elevation. Furthermore, rough high areas are differentiated,

Unit	Tone	Texture	Shape	Size	Height	Location
1	black					
2	grey	smooth				
3	grey	rough			high	mountains
4	grey	rough			low	
5	grey	rough			high	sea + river
6	grey		field			
	white		line			
7	white		field			
	grey		line			
8	grey + black		field	square		
9	grey + black		field	rectangle 5 X 5		
10	grey + black		field	rectangle 20 x 20		

Table 7.1: Made-up example of an interpretation legend.

depending on location in the mountains or along the rivers and the sea.

When delineating areas by hand, there is a limit to what can still be drawn. In practice, polygons smaller than 5 mm by 5 mm should not be drawn. This is called the smallest allowable unit. The scale of the used image(s) thus limits the interpretation cell on the ground. When delineating areas using on-screen digitizing we can zoom in - in principle to a monitor dot. However, we should define the maximum viewing scale at which the given remote sensing data are still reliable. You can then calculate the smallest allowable unit.

Interpretation cell

In some cases an area may consist of two or three different types of too small areas. Individual polygons for each small area can then not be drawn, although

Complex unit

at a larger scale the individual features could be mapped. The solution in such a case is to combine these areas to a complex unit. The different features of such a complex can be described separately. In Table 7.1, Unit 6 and Unit 7 are two different complex units. In Unit 6 there are two features, namely grey fields and white lines, while in Unit 7 there are white fields and grey lines.

7.3.2 Fieldwork

Maps and inventories should reflect what is actually on the ground. Therefore, field visits should be made to observe what is there in reality. Field visits for ground observation are time consuming and usually costly. Making observations everywhere in the entire area to be mapped is likely to take too much time. For efficiency reasons remote sensing data are used to extrapolate the results of a limited amount of observations over the entire area at study.

The selection of sample locations is a crucial step to make mapping cost effective. We can use the RS images to stratify the area. To this end we make an interpretation of the area to be mapped based on the interpretation elements. The interpretation units are the strata to be sampled. In all strata an equal amount of samples is taken. This way of sampling is called *stratified sampling*. We can select the samples in such away that they are representative for the interpretation elements of that unit. This is called *stratified representative sampling*. Stratified representative sampling is a very time-efficient and cost-effective method as compared to random or systematic sampling ([36]; [8]). If an interpretation unit was occupying a very small area, many samples would be needed in the case of random or systematic sampling to make sure that small units are also sampled. When applying the stratified sampling approach, much less samples are needed.

Stratified sampling

Stratified representative sampling can only be applied if the data to be mapped are qualitative (*ie*, nominal or ordinal). For mapping of quantitative data (*ie*, interval or ratio data), *unbiased sampling* strategies (*ie*, random or systematic sampling) should be applied to allow statistical analysis. An example of quantita-

Unbiased sampling

tive data are biomass measurements. Then the entire area needs to be sampled and no prior interpretation is needed for the sampling strategy. Both the stratified and unbiased sampling strategies will be used if quantitative data of certain strata are not required . For instance, we use *stratified random sampling* of grass biomass for livestock management if in the strata forest, water, and urban areas no biomass measurements are needed. We do so to limit the time consuming unbiased sampling.

Field completion

During fieldwork also the location of boundaries of the interpretation is verified. In addition data is gathered about areas or features that cannot be derived from the remote sensing image.

7.3.3 Analyzing field data and map preparation

Based on the correlation between the collected field data and the interpretation, the entire area can be mapped in terms of what is on the ground. If there is a good correlation, only recoding and renaming of the units is required. If the correlation is poor, a complete re-interpretation might be needed after carefully restudying the legend in terms of interpretation elements. For producing the final map, all aspects of map design and cartographic finishing should be observed (see the textbook *Principles of GIS*).

7.4 Quality aspects

The quality of the result of an image interpretation depends on a number of factors: the interpreter, the images 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 in the interpretation.

  Trained interpreter

- The images used limit the phenomena that can be studied, both in a thematic and geometric sense. One cannot, for example, generate a reliable database on the tertiary road system using data from low resolution multispectral scanners. On the other hand, black–and–white aerial photos contain limited information about agricultural crops.

  Adquate images

- 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 should build a seamless database of consistent quality.

  Clear guidelines

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 instants.

Figure 7.4: Two interpretation results derived by two photo-interpreters analyzing the same image. Note the overall differences but also differences in the generalization of the lines. (From [18].)

Replicability

Replicability does not provide information on the accuracy (the relation with the real world) but it does give an indication of the quality of the definition of the features of interest (crisp or ambiguous) and the instructions and methods used. Two examples are given here to give you an intuitive idea. Figure 7.4 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 7.5 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 can also conclude that for some objects (map units) there was no agreement on the thematic attribute.

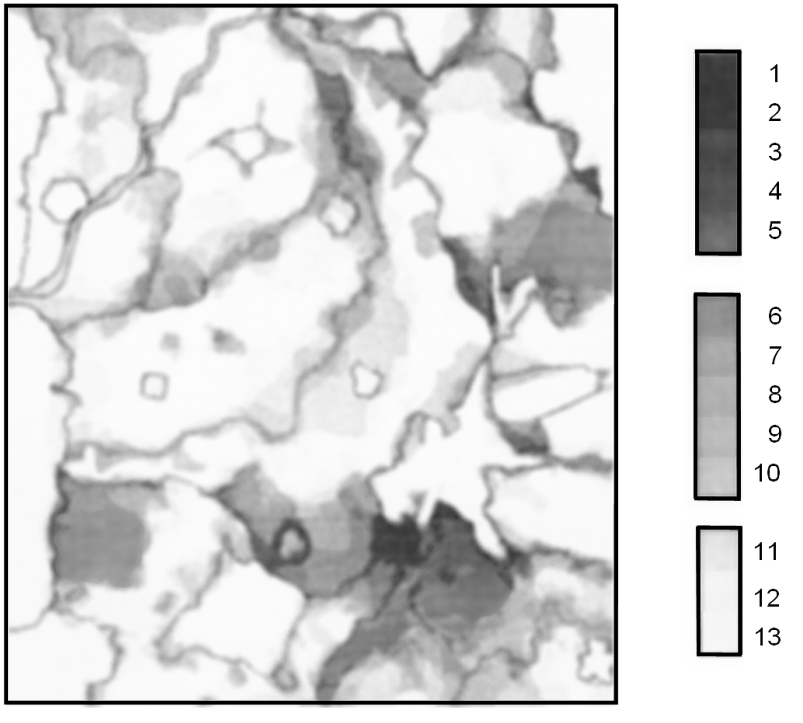


Figure 7.5: Comparison of 13 interpretations of the same image. The grey value represents the the degree of correspondence: white indicates agreement of all 13 interpreters, black indicates that all 13 interpreters disagreed on the thematic class for that location. (From [18].)

Summary

Visual image interpretation is one of the methods used to extract information from remote sensing data. For that purpose, the data need to be visualized on a computer screen or in hardcopy form. Using the human vision system we can distinguish spontaneous recognition and logical inference (reasoning).

Interpretation keys or guidelines are required to instruct the image interpreter. Even if knowledge of the area is lacking, an interpretation legend can be constructed based on interpretation elements, without considering what the units are in reality. After fieldwork it will become clear what the units represent on the ground. In interpretation 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.

The general mapping methodology consists of a systematic interpretation, field work and analysis, and map preparation. In all interpretation and mapping processes, the use of ground observations is essential to acquire knowledge of the local situation, gather data for areas that cannot be mapped from the images, and 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 images available,

and the quality of the guidelines being used.

Additional reading

[15], [22].

Questions

The following questions can help you to study Chapter 7.

1. What is the relationship between data visualization and image interpretation?
2. Compare the list of the interpretation elements in this book with the list in [15] and [22] and describe the difference.
3. Why is an interpretation needed before the field survey in the general mapping approach?
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?

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. What are the steps in the general mapping approach using remote sensing?

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 images.

Chapter 8

Digital image classification

8.1 Introduction

Chapter 7 explained the process of visual image interpretation. In this process, human vision plays a crucial role in extracting information from images. 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.

Digital image interpretation

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 the Food and Agriculture Organization, FAO), digital 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.

Regional scale projects

Image classification is based on the different spectral characteristics of different materials of the Earth's surface, as introduced in Section 2.4. This chapter focuses on the classification of multispectral data. Section 8.2 explains the concepts of image space and feature space. Image classification is a process that operates in feature space. Section 8.3 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 8.4). Section 8.5 discusses problems of standard classification and introduces object oriented classification.

8.2 Principle of image classification

8.2.1 Image space

A digital image is a 2D-array of pixels. The value of a pixel - the DN - is in the case of 8 bits recording in the range 0 to 255. A DN corresponds to the energy reflected or emitted from a ground resolution cell unless the image has been resampled. The spatial distribution of the DNs defines the image or *image space*. A multispectral sensor records the radiation from a particular GRC in different channels according to its spectral band separation. A sensor recording in three bands (Figure 2.24) yields three pixels with the same row-column tuple (i, j) stemming from one and the same GRC.

8.2.2 Feature space

When we consider a two-band image, we can say that the 2 DNs for a GRC are components of a two-dimensional vector $[v1, v2]$, the *feature vector* (see Figure 8.1). An example of a feature vector is $[13, 55]$, which tells that the conjugate pixels of band 1 and band 2 have the DNs 13 and 55. This vector can be plotted in a two-dimensional graph.

Feature vector

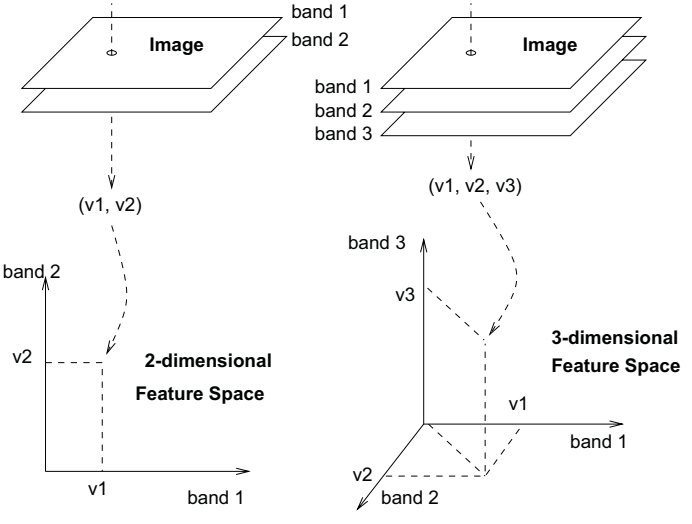


Figure 8.1: Plotting of the pixel values of a GRC in the feature space for a two-band and three-band image.

Similarly we can visualize a three-dimensional feature vector $[v1, v2, v3]$ of a cell in a three-band image in a three-dimensional graph. A graph that shows the feature vectors is called a *feature space*, or 'feature space plot' or 'scatter plot'. Figure 8.1 illustrates how a feature vector (related to one GRC) is plotted in the

feature space for two and three bands, respectively. Two-dimensional feature space plots are most common.

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

Scatterplot

Plotting all the feature vectors of digital image pair yields a 2D scatterplot of many points (Figure 8.2). A 2D scatterplot provides information about pixel value pairs that occur within a two-band image. Note that some combinations will occur more frequently, which can be visualized by using intensity or colour (as introduced in Chapter 5).

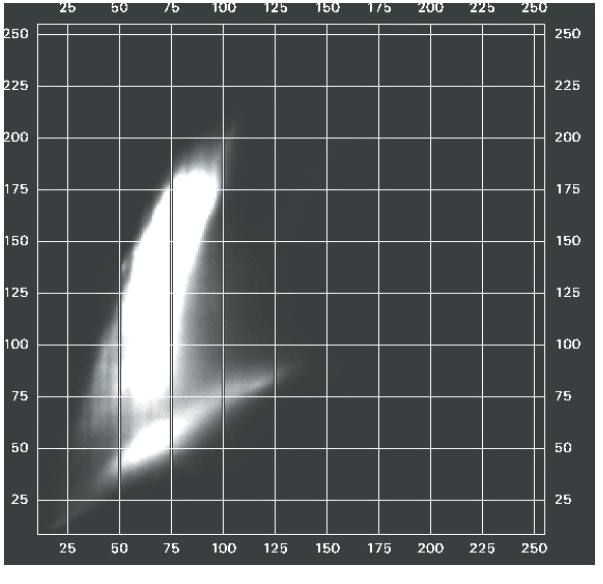


Figure 8.2: Scatterplot of two bands of a RS image. Note the units along the x- and y-axes. The intensity at a point in the feature space is related to the number of cells at that point.

Distances and clusters in the feature space

We use distance in the feature space to accomplish classification. Distance in the feature space is measured as 'Euclidian distance' in the same unit as the DNs (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 8.3, 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.

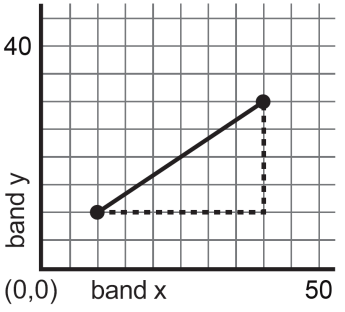


Figure 8.3: Euclidian distance between the two points is calculated using Pythagoras' theorem.

8.2.3 Image classification

The scatterplot shown in Figure 8.2 shows the distribution of conjugate pixel values of an actual two-band image. Figure 8.4 shows a feature space in which the feature vectors have been plotted of samples of six specific land cover classes (grass, water, trees, etc). You can see that the feature vectors of GRCs that are water areas form a compact cluster. Also the feature vectors of the of other land cover types (classes) are clustered. Most of these clusters occupy their own area in the feature space, only the grass and wheat clusters overlap to a larger extent. Figure 8.4 illustrates 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 feature vector of a multi-band image can be plotted and checked against these classes and assigned to the class where it fits best.

Cluster

Classes

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 (see Section 2.4). Figure 8.4 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.

Spectral differentiation

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 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

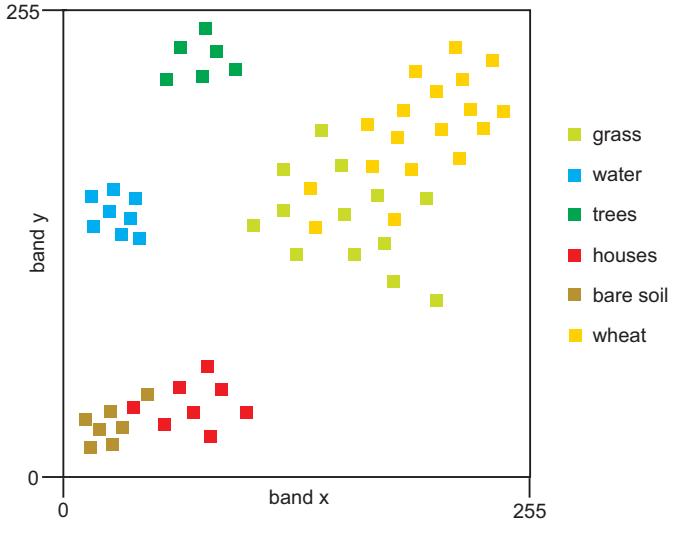


Figure 8.4: Feature space showing the respective clusters of six classes; note that each class occupies a limited area in the feature space.

process and is carried out during the 'training process'. Comparison of the individual pixels with the clusters takes place using 'classification algorithms'. Both are explained in the next section.

8.3 Image classification process

The process of image classification (Figure 8.5) typically involves five steps:

1. Selection and preparation of the RS images. Depending on the land cover types or whatever needs to be classified, the most appropriate sensor, the most appropriate date(s) of acquisition and the most appropriate wavelength bands should be selected (Section 8.3.1).

Data selection

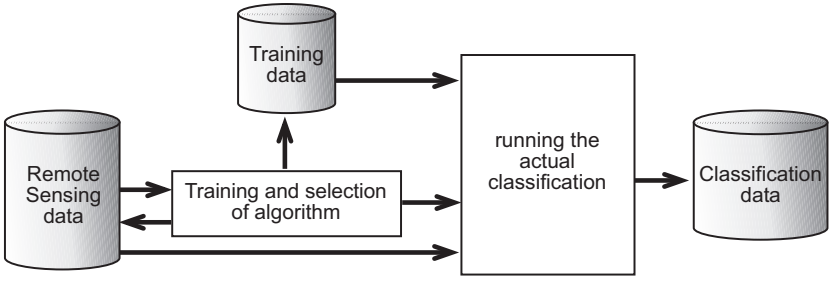


Figure 8.5: The classification process; the most important component is training in combination with selection of the algorithm.

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 8.3.2); in an unsupervised classification a clustering algorithm automatically finds and defines a number of clusters in the feature space (Section 8.3.3).

Supervised or unsupervised

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

Algorithm selection

(based on their feature vectors) are assigned to the classes. The assignment can be based on different criteria (Section 8.3.4).

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 DNs, each "multi-band pixel" (cell) in the image is assigned to one of the predefined classes (Figure 8.6). Classifying

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 8.4). Validating

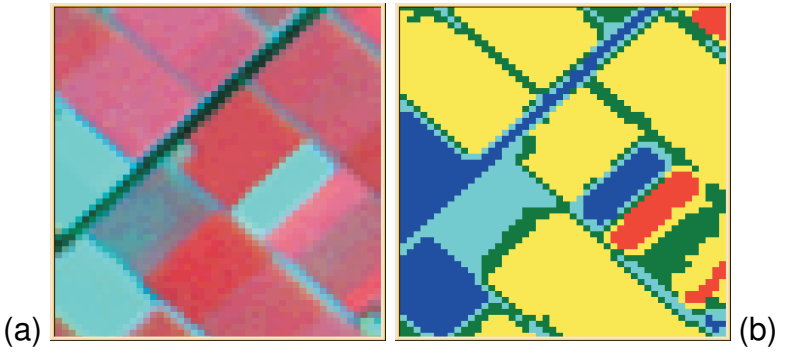


Figure 8.6: The result of classification of a multi-band image (a) is a raster in which each cell is assigned to some thematic class (b).

The above points are elaborated on 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, however, limits itself to an image that is composed of a maximum of three bands.

8.3.1 Preparation for image classification

Image classification serves a specific goal: converting RS images to thematic data. In the application context, one is rather interested in thematic characteristics of a scene 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 into models. In addition, image classification can also be considered as data reduction: the n multispectral bands result in a single-band image file.

Thematic classes

With a particular application in mind, the information classes of interest need to be defined. 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 images 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.

Data date

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-5 TM) lie in the problem of band correlation and, sometimes, in limitations of hardware 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

Selection of bands

classification process.

8.3.2 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 is familiar with the area of interest. The operator needs to know where to find the classes of interest in the scene. This information can be derived from general knowledge of the scene or from dedicated field observations.

Scene knowledge

A sample of a specific class, comprising a number of training cells , forms a cluster in the feature space (as portrayed in Figure 8.4). The clusters, as selected by the operator

Training set

- 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 an 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 only partially overlap with the other clusters, as otherwise a reliable separation is not possible. Using a specific multi-band image, 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 images acquired at other moments.

The resulting clusters can be characterized by simple statistics of the point distributions. These are for one cluster: the vector of mean values of the DNs (for band 1 and band 2, see Figure 8.9), and the standard deviations of the DNs (for band 1 and band 2, see Figure 8.10, where the standard deviations are plotted as crosses).

8.3.3 Unsupervised image classification

Supervised classification requires knowledge of the area at hand. If this knowledge is not sufficiently 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 exist, their main purpose being to produce spectral groupings based on certain spectral similarities. In one of the most common approaches, the user has to define the maximum number of clusters for an image at hand. Based on this, the computer locates arbitrary mean vectors as the centre points of the clusters. Each cell is then assigned to a cluster by the 'minimum distance to cluster centroid' decision rule. Once all the cells have been labelled, recalculation of the cluster centre takes place and the process is repeated until the proper cluster centres are found and the cells 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 cells 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 measures. Merging of clusters needs to be done to reduce the number of unnecessary subdivisions of the feature space. 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 cells as a threshold for cluster elimination. Analysis of the cluster compactness around its centre point is done by

Number of classes

Iteration process

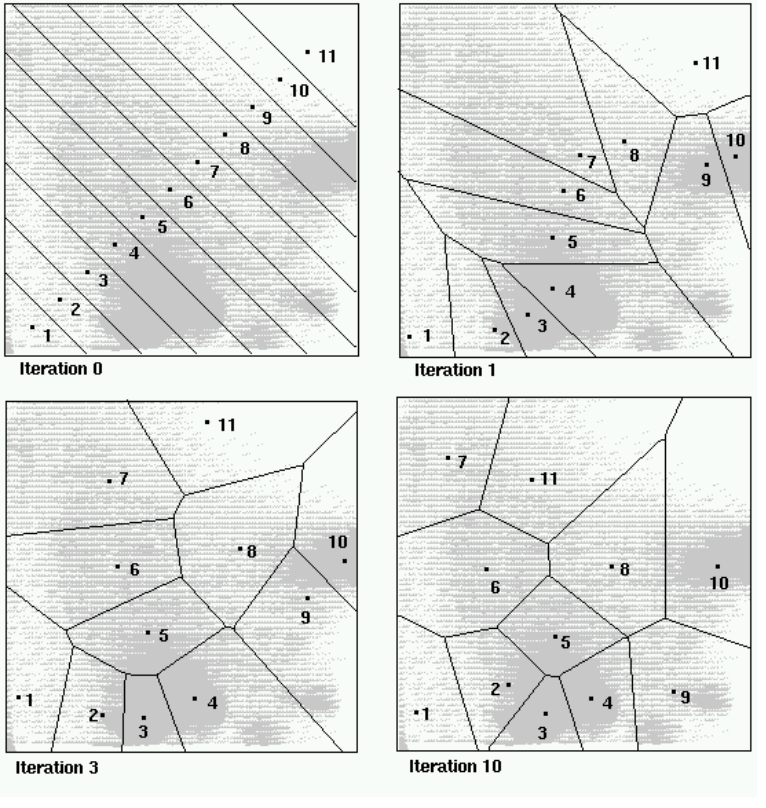


Figure 8.7: The subsequent results of an iterative clustering algorithm.

means of the user-defined standard deviation for each spectral band. If a cluster is elongated, separation of the cluster will be done perpendicular 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. The clusters that result after the last iteration are described by their statistics. Figure 8.7 shows the results of a clustering algorithm for an example image. 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.

8.3.4 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, 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 algorithms are explained. First the *box classifier* is explained, for its simplicity to help in 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 most frequently.

Box classifier

The box classifier is the simplest 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 (Figure 8.8). The number of boxes depends on the number of classes. During classification, every feature vector of an input (two-band) image will be checked to see if it falls in any of the boxes. If so, the cell will get the class label of the box it belongs to. Cells that do not fall inside any of the boxes will be assigned the 'unknown class', sometimes also referred to as the 'reject class'.

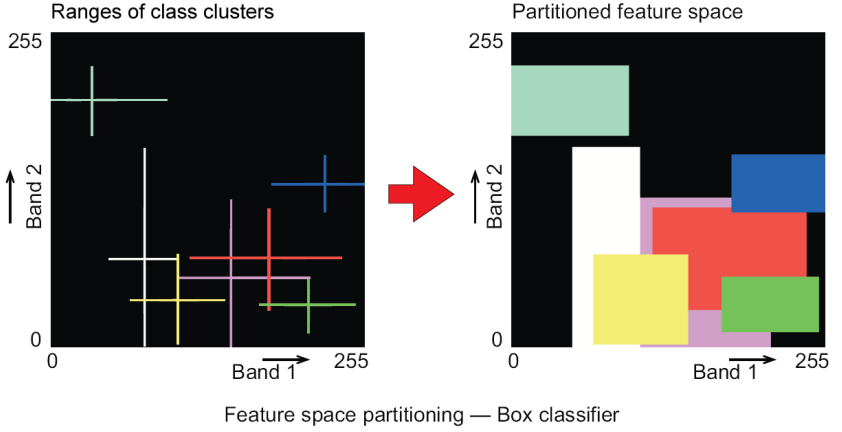


Figure 8.8: Principle of the box classification in a two-dimensional situation.

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

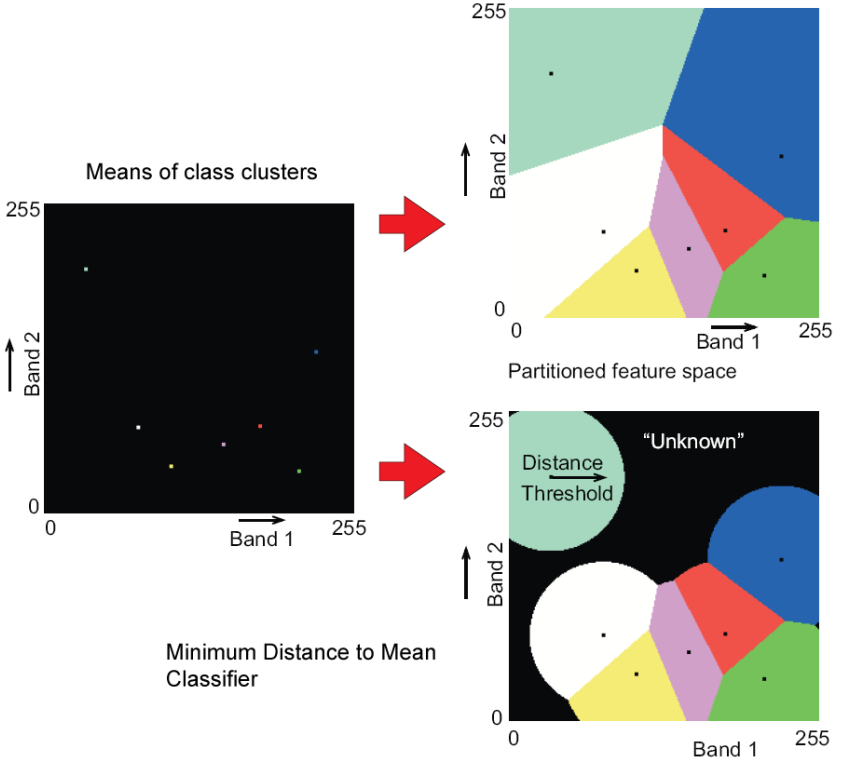


Figure 8.9: Principle of the minimum distance to mean classification in a two-dimensional situation. The decision boundaries are shown for a situation without threshold distance (upper right) and with threshold distance (lower right).

Minimum Distance to Mean classifier

The basis for the Minimum Distance to Mean (MDM) classifier are the cluster centres. During classification the Euclidean distances from a candidate feature vector to all the cluster centres are calculated. The candidate cell is assigned to the class which qualifies as the closet one. Figure 8.9 illustrates how a feature space is partitioned based on the cluster centres. One of the disadvantages of the MDM classifier is that points 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 8.9 illustrates the 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.

Maximum Likelihood classifier

The Maximum Likelihood (ML) classifier considers not only the cluster centres but also the shape, size and orientation of the clusters. 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. A cell 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 follow 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 8.10 shows the decision boundaries for a situation with and without threshold distance.

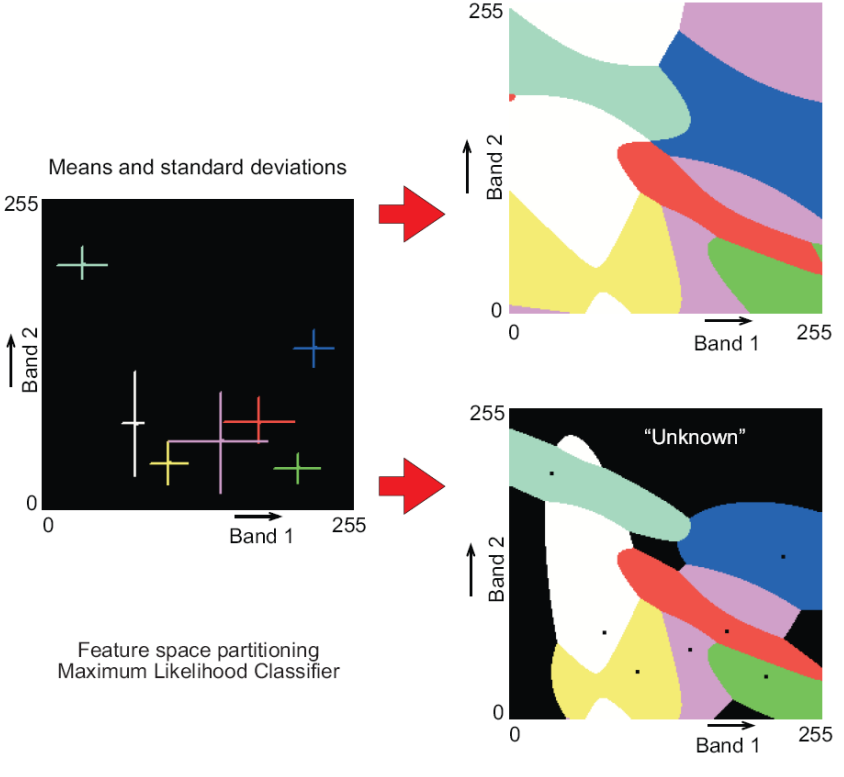


Figure 8.10: Principle of the maximum likelihood classification. The decision boundaries are shown for a situation without threshold distance (upper right) and with threshold distance (lower right).

8.4 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 of the result should be checked. This is usually done by a sampling approach in which a number of raster elements of the output 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' is preferably derived from field observations. Sometimes sources of an assumed higher accuracy, such as aerial photos, are used as a reference.

Sampling schemes

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.

	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 8.1: 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%.

Error matrix

Once the sampling has been carried out, an error matrix can be established (Table 8.1). Other terms for this table are *confusion matrix* or *contingency matrix*. In the table, four classes (A, B, C, D) are listed. A total 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.

Overall accuracy

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 (*ie*, the sum of the diagonal cells in the error matrix) divided by the total number of pixels checked. In Table 8.1 the overall accuracy is $(35 + 11 + 38 + 2)/163 = 53\%$. The overall accuracy yields one figure

for the classification 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.

Omission

Commission

Omission error is the corollary of producer accuracy, while user accuracy is the corollary of commission error. 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.

Producer accuracy

User accuracy

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 different accuracy. This type of testing is used to evaluate different RS data or methods for the generation of spatial data.

Kappa

8.5 Pixel-based and object oriented classification

Pixel-based image classification is a powerful technique to derive 'thematic classes' from multi-band images. 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

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 to land cover classes. 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 comprise several land cover classes. Table 8.2 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 images (other bands, other moments) or existing geospatial data, such as topographic maps, historical land inventories, road maps, etc. Usually this is done in combination with adding expert 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 8.2: 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 pixel-based image classification is that each pixel is only assigned to one class. This is not a problem when dealing with (relatively) small ground resolution cells. However, when dealing with (relatively) large GRCs, more land cover classes are likely to occur within a cell. As a result, the value of the pixel is an average of the reflectance of the land cover present within the GRC. 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 8.11). 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

Mixed pixels

using data with the appropriate spatial resolution.

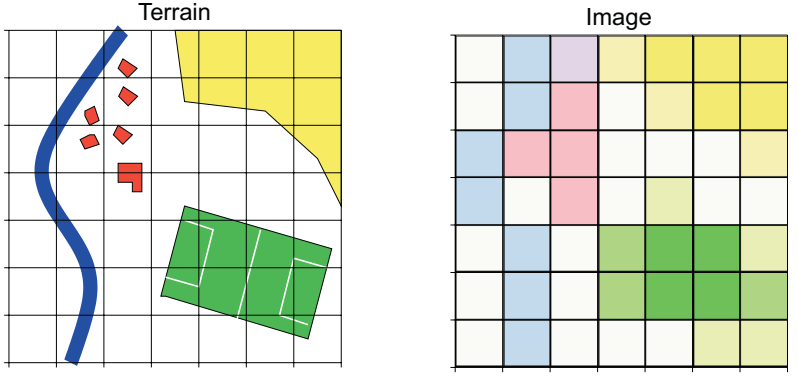


Figure 8.11: The origin of mixed pixels: different land cover types occur within one ground resolution cell. Note the relative abundance of mixed pixels.

In the previous sections you learnt that the choice of the classification approach depends on the data available and also the knowledge we have about the area under investigation. Without knowledge on the present land cover classes, unsupervised classification can give an overview of the variety of classes in an image. If knowledge is available, such as from field work or other sources, supervised classification may be superior. However, both methods only make use of spectral information, which gets increasingly problematic with higher spatial resolution. For example a building that is made up of different materials leads to pixels with highly variable spectral characteristics, and thus a situation where training pixels are of little help. Similarly, a field may contain healthy vegetation pixels as well as some of bare soil. We are also increasingly interested in land use. However, to distinguish for example urban from rural woodland, or a swimming pool from a natural pond, an approach similar the visual interpretation as described in Chapter 7 is needed. Object-oriented analysis (OOA),

Object-oriented analysis

also called segmentation-based analysis, allows us to do that. Instead of trying to classify every pixel separately and only based on spectral information, OOA breaks down an image into spectrally homogenous segments that correspond to fields, tree stands, buildings, etc. It is also possible to use auxiliary GIS layers, for example building footprints, to guide this segmentation. Similar to the cognitive approach of visual image interpretation - where we consider each element in terms of its spectral appearance but also in terms of its shape and texture, and within its environment - in OOA we can then specify contextual relationships and more complex segment characteristics to classify the objects extracted in the segmentation process. For example, we can use object texture to distinguish two spectrally similar forest types, or distinguish a swimming pool from a pond by considering its shape and perhaps the surrounding concrete in place of soil and vegetation. OOA is particularly suitable for images of high spatial resolution, but also for data obtained by ALS or microwave radar. It requires that we have substantial knowledge on what distinguishes a given land cover or land use type, as well as auxiliary data such as elevation, soil type, or vector layers.

Summary

Digital image classification is a technique to derive thematic classes from RS images. Input is a multi-band image. Output is a raster file containing thematic (nominal) classes. In the process of image classification the role of the operator and additional (field) data are significant. In a supervised classification, 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.

Additional reading

[15], [19].

Questions

The following questions can help you to study Chapter 8.

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. Name 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.

Capita selecta

Chapter 9

Aerial photography

9.1 Introduction

Aerial photographs have been used since the early 20^{th} century to provide geospatial data for a wide range of applications. *Photography* is the process or art of producing images by light on a sensitive surface. Taking and using photographs is the oldest, yet most commonly applied remote sensing technique. *Photogrammetry* is the science and technique of making measurements on photos and converting these to quantities which are meaningful in the terrain. ITC started out on photography and photogrammetry, the latter being at that time the most innovative and promising technique of topographic mapping of large areas. Aerial film cameras are typically mounted in an aircraft, however, there was a Russian satellite carrying a photographic camera and a NASA Space Shuttle mission taking systematic photo coverage with a German camera. Aerial photos and their digital variant obtained by digital frame cameras are today the prime data source for medium to large scale topographic mapping, for many cadastral surveys, civil engineering projects, and urban planning. Aerial photographs are also a useful source of information for foresters, ecologists, soil scientists, geologists, and many others. Photographic film is a very mature medium and aerial survey cameras using film have reached vast operational maturity in the course of many years, hence new significant developments cannot be expected. Owners of aerial film cameras will continue using them as long as Agfa and Kodak continue producing films at affordable prices.

Two broad categories of aerial photographs can be distinguished: 'vertical' and 'oblique' photographs (Figure 9.1). In most mapping applications, vertical aerial photographs are required. A vertical aerial photograph is produced with a cam-

era mounted in the floor of an aircraft. The resulting image is similar to a map and has a scale that is approximately constant throughout the image area. Vertical aerial photographs for mapping are usually taken such that they overlap in the flight direction by at least 60 %. Two successive photos can form a stereo pair, thus enabling 3D measurements.

Vertical photo

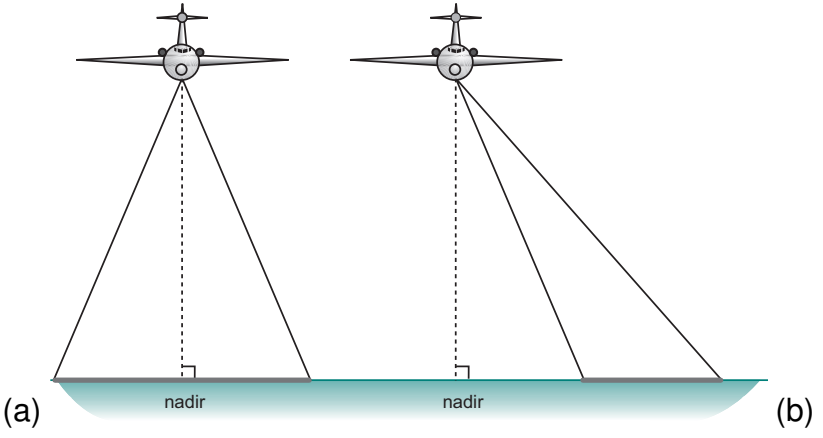


Figure 9.1: Vertical (a) and oblique (b) photography.

Oblique photographs are obtained when the axis of the camera is not vertical. They can also 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.

Oblique photo

(a)   (b)

Figure 9.2: Vertical (a) and oblique (b) aerial photo of the ITC building, 1999.

This chapter focuses on the camera, films and methods used for vertical aerial photography. Section 9.2 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 image products (Section 9.3). The Section 9.4 treats basic geometric characteristics of aerial photographs. In Section 9.5 the basics of aerial photography missions are introduced. The Section 9.6 gives a brief account on scanning aerial photographs.

9.2 Aerial survey camera

A camera used for vertical aerial photography for mapping purposes is called an *aerial survey camera*. Only two manufacturers of aerial survey cameras, namely Leica and Z/I Imaging have continued to assemble aerial film cameras until recently. Their cameras are the RC-30 and the RMK-TOP, 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. 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 very large film size. Modern aerial survey cameras produce negatives measuring 23 cm $\times$ 23 cm (9 by 9 inch). Up to 600 photographs may be recorded on a single roll of film. For the digital camera to have the same kind of quality as the aerial film camera it would have to have about 200 million pixels.

9.2.1 Lens cone

The most expensive component of an aerial survey camera is the lens cone. It is interchangeable and the manufacturers produce a range of cones, each of different focal length. The focal length together with the flying height determine the photo scale (Section 9.4.1). 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.

Wide angle lens

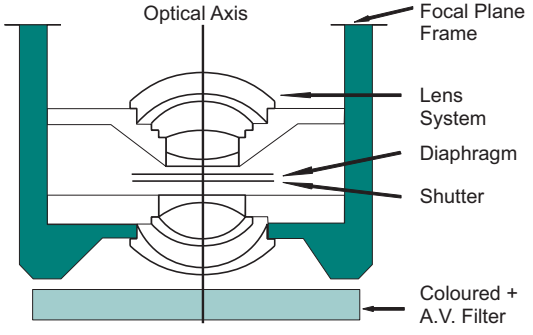


Figure 9.3: The lens cone comprises a lens system, focal plane frame, shutter, diaphragm, anti-vignetting and coloured filters.

The lens is built such that a best possible approximation of a *central projection* is achieved, *ie*, that we can geometrically model the object-image relationship as central perspective: all light rays passing through one point, the projection centre. Lenses for survey cameras achieve a projection to the focal plane which differs from an ideal central projection only by a few microns. This difference is referred to as *lens distortion*. For economic reasons photogrammetrists are of-

Lens distortion

ten interested in measuring very precisely in images, determining image coordinates with a precision better than 5 μm. To this end lens distortion is measured on a regular basis in specialized laboratories and reported in a 'calibration report'. This report is required in many photogrammetric processes.

In the lens cone (Figure 9.3), the diaphragm and shutter, respectively, control the intensity and the duration of light 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 for, normally leads to a reduction in image contrast.
- 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; it is related to the angle with the optical axis. The greater the angle, the darker the image. In wide angle lenses, this effect can be severe. 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.

9.2.2 Film magazine and auxiliary data

The aerial camera is fitted with a system to record various items of relevant information onto the side of the negative: mission identifier, date and time, flying height and the frame number (Figure 9.4).

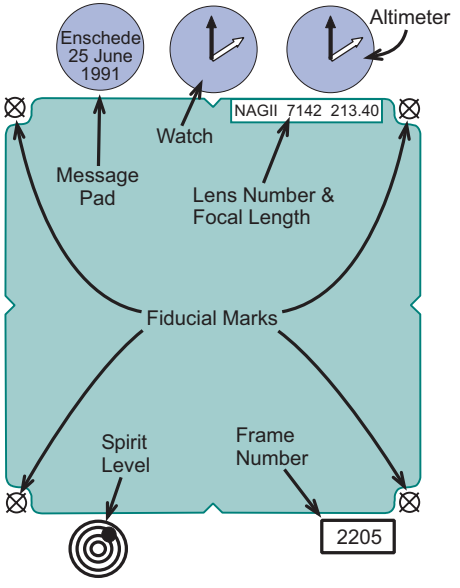


Figure 9.4: Auxiliary data annotation on an aerial photograph.

A vacuum plate is used for flattening the film at the instant of exposure. So-called 'fiducial marks' are recorded in all corners and/or on the sides of the film. The fiducial marks are required to determine the optical centre of the photo, which in turn is needed to relate measurements made in the photo to position in

Fiducial marks

the terrain (see Chapter 6).

Film magazines of modern cameras are equipped with an image motion compensation device. *Forward motion compensation* in aerial survey cameras causes the image projected on the film to be displaced across the film during the time that the shutter is open. Forward motion is potentially the most damaging of the effects that disturb the geometric quality of the photograph. It 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.

Forward motion compensation

9.3 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.

We have two main types of film: black-and-white (B&W) film and colour film. A B&W film, such as a panchromatic film, has one emulsion layer, colour films have three emulsion layers (Figure 9.5). 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*.

Emulsion

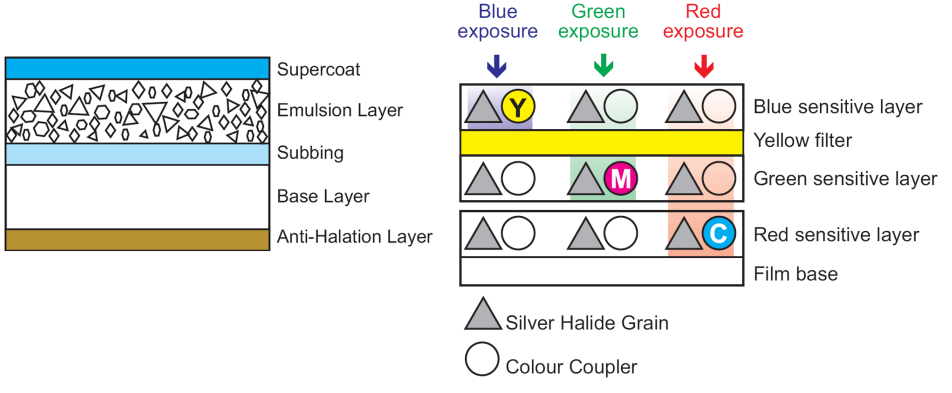


Figure 9.5: Layers of B&W and colour film.

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

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

In the following section the general sensitivity is explained, followed by an explanation of the spectral sensitivity. Subsequently, colour and colour infrared photographs are explained.

9.3.1 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.

Grain sensitivity

The sensitivity of a film can be increased by using larger silver halide 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, *ie*, the spatial resolution is decreased (Section 9.4.2). An alternative to improve the general sensitivity of a film is to apply a sensitization of the emulsion by adding small quantities of chemicals, such as gold or sulphur.

Film speed

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 latent image. Similarly, the higher the film speed, the less bright an object needs to be in order to be recorded.

9.3.2 Spectral sensitivity

Sensitization techniques are used not only to increase the general sensitivity but also to produce films that are sensitive to light of longer wavelengths than those of blue light. By adding sensitizing dyes to the basic silver halide emulsion a film can be made sensitive to green, red or infrared waves. We can produce different B&W films by using different sensitization. Most common are the panchromatic and infrared sensitive film. The sensitivity curves of these films are shown in Figure 9.6.

Sensitizing dyes

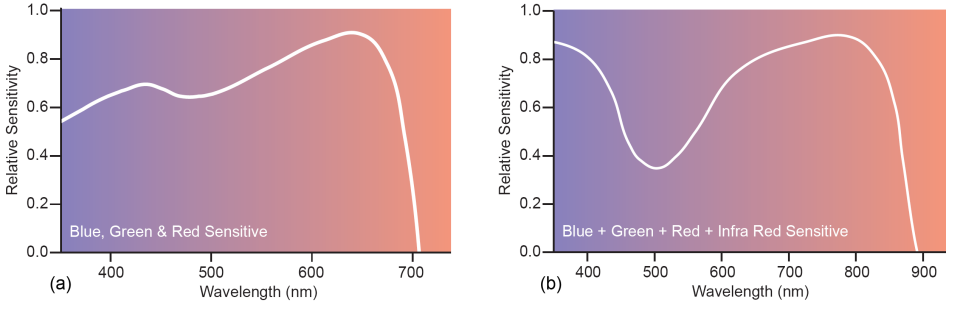


Figure 9.6: Spectral sensitivity curves of a panchromatic film and a B&W IR film. Note the difference in scaling on the x-axis.

9.3.3 True colour and colour infrared photography

We need a film with three light sensitive layers to obtain colour photos. For RS we use two types of colour film, 'true colour' and 'false colour IR' film. Figure 9.7 shows examples of panchromatic and NIR-B&W images and of true and false colour images.

In colour photography, each of the primary colours creates colour particles of the "opposite" colour, *ie*, dyes that remove this colour, but allow the other two colours to pass. The result is a colour negative. If we make a colour copy of that negative, we will get the original colours in that copy.

True colour

The emulsion used for 'colour infrared film' creates yellow dyes for green light, magenta dyes for red light, and cyan dyes for IR. Blue light should be kept out of the camera by a filter. IR, Red and Green give the same result as Red, Green and Blue, respectively, in the normal case. If a copy of this 'IR-negative' is made with a normal colour emulsion, the result is an image which shows blue for green objects, green for red objects and red for IR objects. This is called a *false colour IR* image.

Colour infrared

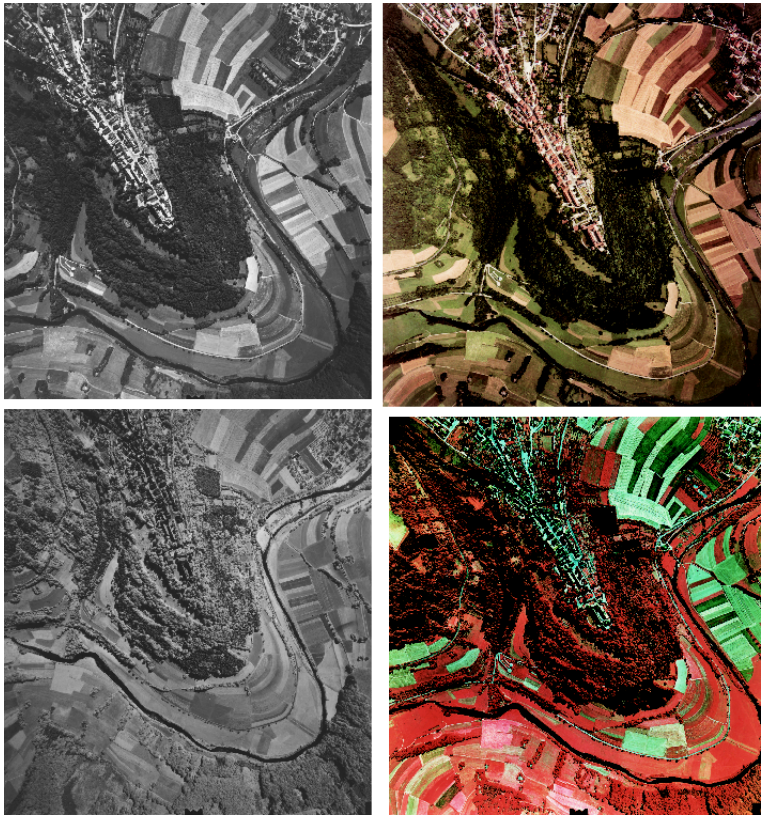


Figure 9.7: Single-band and three-band photos in the visible and NIR part of the EM spectrum.

9.4 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 for different focal lengths.

9.4.1 Scale

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

$$s = \frac{H}{f}. \tag{9.1}$$

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, while the flying height remains constant, then:

- the photo scale factor will increase and the size of the individual details in the image becomes smaller. In the example shown in Figure 9.8, using a 150 mm and 300 mm lens at H =2000 m results in a scale factor of 13,333 and 6,666, respectively;
- the ground coverage increases. A 23 cm negative covers an area of 3066 m by 3066 m if f = 150 mm. The width of the area reduces to 1533 m if f = 300 mm. Subsequent processing takes less time if we can cover a large area by less photos;

Ground coverage

- the FOV increases and the image perspective changes. The FOV for the wide angle lens is 74°, for the normal angle lens (300 mm) it is 41°. Using shorter focal lengths has the advantage of achieving more precise elevation measurements in stereo images (see Chapter 6). Flying a camera with a wide angle lens at low altitudes creates the disadvantage of larger obscured areas: if there are tall buildings near the edges of a photographed

Field of view

scene the areas behind the buildings become hidden because of the central perspective; we call it the *dead ground effect*.

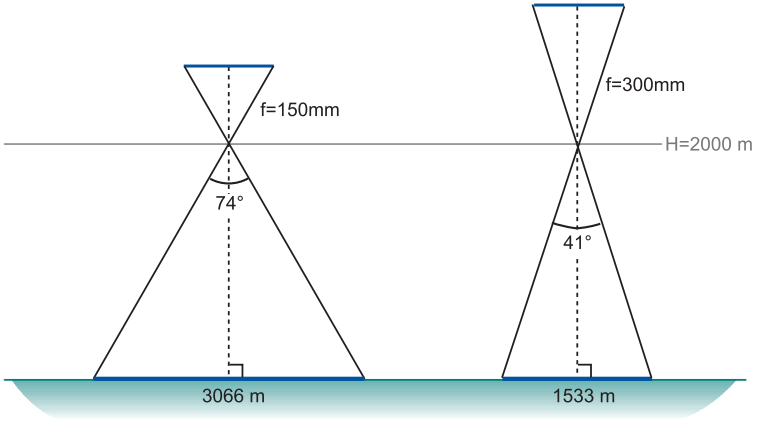


Figure 9.8: Effect of a different focal length at the same flying height on ground coverage.

9.4.2 Spatial resolution

Whereas 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 distinguish small adjacent objects in an image. The spatial resolution of B&W 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:

Lines per mm

- 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 even during a mission.

9.5 Aerial photography missions

Mission planning If a mapping project requires aerial photographs, one of the first tasks is to select the required photo scale, the type of lens to be used, the type of film to be used, and the required percentage of overlap (for stereo). Forward overlap usually is around 60%, while sideways overlap typically is around 20%. Figure 9.9 shows a survey area that is covered by a number of flight lines. Furthermore, the date and time of acquisition should be considered with respect to growing season, light conditions and shadowing effects.

Overlap

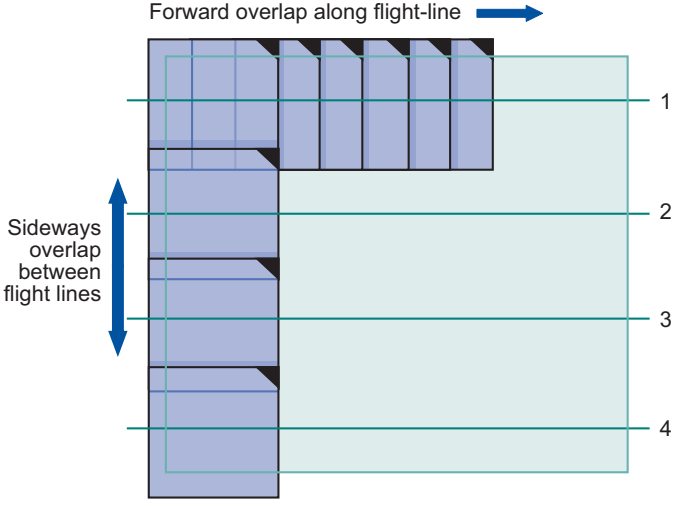


Figure 9.9: Arrangement of photos in a typical 'aerial photo block'.

Once 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, either mission maps are prepared for use by the survey navigator in the case of a conventional mission execution, or otherwise the data are input to the mission guidance system.

Mission execution In current professional practise, we use a computer program to determine - after entering a number of relevant mission parameters and the area of interest - the (3D) coordinates of all positions from which photographs are to be taken. These are stored in a job database. On board, the camera operator/pilot can obtain all relevant information from that database, such as project area, camera and type of film to be used, and number of images, constraints regarding time of day or sun angle, season, and atmospheric conditions.

The camera positions being loaded to a guidance system, the pilot is guided along the flight lines – with the support of GPS – such that the deviation from the ideal line (horizontal and vertical) and the time to the next exposure station is shown on a display (together with other relevant parameters). If the airplane passes 'close enough' to the predetermined exposure station, the camera is fired automatically at the nearest position. This allows having the data of several projects on board and making the choice of project (or project part) according to

the local weather conditions. If necessary, one can also abandon a project and resume it later.

In the absence of GPS-guidance the navigator has to observe the terrain using the traditional viewing system of an aerial camera, check the flight lines against the planned ones, which are shown graphically in topographic maps, and give the required corrections to the left or to the right to the pilot, and also to tune the overlap regulator to the apparent forward speed of the airplane.

Global navigation satellite systems

Global navigation satellite systems (GPS-USA, Glonass-SU, and the forthcoming 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. Computer controlled navigation and camera management is especially important in surveying areas where topographic maps do not exist, are old, or are of small scale or poor quality. It is also helpful in areas where the terrain has few features (sand deserts, dense 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.

9.6 Scanning photographs

Classical photogrammetric techniques as well as visual photo-interpretation generally employ hardcopy photographic images. These can be the original negatives, positive prints or diapositives. Digital photogrammetric systems, as well as geographic information systems, require digital images. A scanner is used to convert a film or print into a digital form. The scanner samples the image with an electro-optical detector and measures the brightness of small areas (pixels). The brightness values are then represented as a digital number on a given scale. In the case of a B&W image, a single measurement is made for each pixel area. In the case of a colour 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 images are to be used in precise photogrammetric procedures. Today flatbed scanners are the most common type. Scanning is done by moving a linear CCD array along the document.

Flatbed scanners

In the scanning process, the setting of the size of the scanning resolution 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 photo detail the scanned image should reveal as required for the application; it is limited by the scanner and the type of film. Office scanners permit around 600 dpi, which gives a dot size (pixel size) of 42 µm (2.54 cm ÷ 600 dots). Photogrammetric scanners, on the other hand, may produce 3600 dpi (7 µm dot size).

Scanning density

For a B&W 23 cm × 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

File size

used per pixel (*ie*, there are 256 grey levels), the resulting files requires 29 Mbyte of disk space. When the scale of the negative is given, the pixel size on the ground 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 (see Formula 4.1). The pixel size on the ground of the resulting image is, therefore, 0.72 m.

Pixel size on the ground

Summary

This chapter provided an overview of aerial photography. First, the characteristics of oblique and vertical aerial photographs were distinguished. Taking vertical aerial photographs requires a specially adapted aircraft.

The main components of an aerial camera system are the lens and the film. The focal length of the lens, in combination with the flying height, determines the photo scale factor. The film type used determines which wavelength bands are recorded. The most commonly used film types are panchromatic, black-and-white infrared, true-colour and false-colour infrared. Other characteristics of film are the general sensitivity, which is related to the size of the grains, and spectral sensitivity, which is related to the wavelengths that the film is sensitive to. After exposure, the film is developed and printed. The film or printed photo can be scanned for use in a digital environment.

There is professional software for mission planning. For mission execution we rely at least on GPS but increasingly on a computer controlled mission guidance system. A high end aerial survey camera system will include, moreover, a POS for direct camera orientation.

Additional reading

[15], [16], [17], [34].

Questions

The following questions can help you to study Chapter 9.

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, focal length, and overlap?
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, and other parameters you consider relevant).
3. Compare multispectral sensor data with scanned aerial photographs. Which similarities and differences can you identify?

The following are typical exam questions:

1. Calculate the scale factor for an aerial photo taken at a flying height of 2500 m by a camera with a focal length of 88 mm.
2. Consider a black-and-white film. What determines the general sensitivity of this film and why is it important?
3. Explain spectral sensitivity.
4. A hardcopy aerial photograph is to be scanned using a flatbed 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 10

Active sensors

10.1 Introduction

Active remote sensing technologies have the potential to provide accurate information about the terrain surface by imaging SAR from airborne or spaceborne platforms and three-dimensional measurement of the surface by interferometric SAR (InSAR or IfSAR), and airborne laser scanning (ALS). While these technologies are both active ranging systems, they represent fundamentally different sensing processes. Radar systems, introduced in Section 10.2, are based upon microwave sensing principles, while laser scanners are optical sensors, operating in the visible or near-infrared portion of the electromagnetic spectrum. They are introduced in Section 10.3.

10.2 Radar

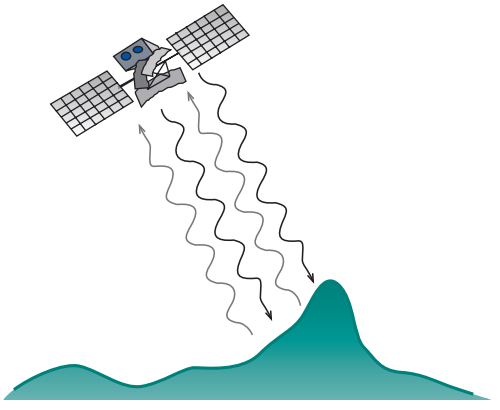


Figure 10.1: Principle of active microwave remote sensing.

10.2.1 What is radar?

Microwave remote sensing uses electromagnetic waves with wavelengths between 1 cm and 1 m (see Figure 2.7 and 10.3). These relatively long wavelengths have the advantage that they can penetrate clouds and are independent of atmospheric scattering. Although microwave remote sensing is primarily considered an active technique, also passive sensors are used. Microwave radiometers operate similarly to thermal sensors by detecting naturally emitted microwave energy (either terrestrial or atmospheric). They are primarily used in meteorology, hydrology and oceanography. In active systems on the other hand, 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 10.1. There are several advantages to be gained from the use of active sensors, which have their own energy source:

Microwave RS

- it is possible to acquire data at any time, also during the night (similar to thermal remote sensing);
- since the waves are created by the sensor itself, the signal characteristics are fully controlled (*eg*, wavelength, polarization, incidence angle, etc) and can be adjusted according to the desired application.

Active sensors can be divided into two groups: imaging and non-imaging sensors. Radar sensors are typically 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 sensors are widely used in civil applications as well, such as environmental monitoring. To the group of non-imaging microwave instruments belong *altimeters*, which collect distance information (*eg*, sea surface elevation), and *scatterometers*, which acquire information about the object properties (*eg*, wind speed).

Non-imaging radar

This section focuses on the principles of imaging radar and its applications. The interpretation of radar images is less intuitive than the interpretation of photographs and similar images. This is because of differences in physical interaction of the waves with the Earth's surface. The following explains which interactions take place and how radar images can be interpreted.

10.2.2 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*:

Backscattered energy

$$P_r = \frac{G^2 \lambda^2 P_t \sigma}{(4\pi)^3 R^4}, \tag{10.1}$$

where

P_r = received energy,
G = antenna gain,
λ = wavelength,
P_t = transmitted energy,
σ = 'radar cross section', which is a function of the object characteristics and the size of the illuminated area, and
R = 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, *ie*, wavelength, antenna and transmitted power,
- radar imaging geometry, that defines the size of the illuminated area which is a function of, for example, beam-width, incidence angle and range, and
- characteristics of interaction of the radar signal with objects, *ie*, surface roughness and composition, and terrain relief (magnitude and orientation of slopes).

They are explained in the following sections in more detail.

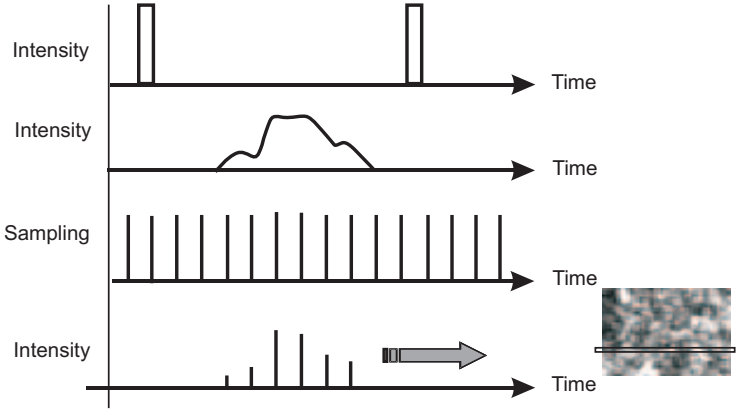


Figure 10.2: Illustration of how radar pixels result from pulses. For each sequence shown, one image line is generated.

What exactly does a radar system measure? To interpret radar images correctly, it is important to understand what a radar sensor detects. The physical properties of a radar wave are the same as those introduced in Section 2.2. Radar waves are electric and magnetic fields that oscillate like a sine wave in perpendicular planes. In dealing with radar, the concepts of wavelength, period, frequency, amplitude, and phase are relevant.

The radar transmitter creates microwave signals, *ie*, 'pulses' of microwaves at a fixed frequency (the *Pulse Repetition Frequency*), that are directed by the antenna into a 'beam'. A pulse travels in this beam 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 is received and its intensity measured. 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 calculated (see Formula 10.2).

Distance

Intensity

To create an 'image', the return signal of a single pulse is sampled and these samples are stored in an image line (Figure 10.2). With the movement of the sensor while emitting pulses, a two-dimensional image is created (each pulse defines one line). The radar sensor, therefore, measures distances and backscattered signal intensities.

Commonly used imaging radar bands Similarly to optical remote sensing, radar sensors operate with one or more different bands. For better identification, a standard has been established that defines various wavelength classes

using letters to distinguish the various bands (Figure 10.3). In the description of different radar missions you will recognize the different wavelengths used when 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.

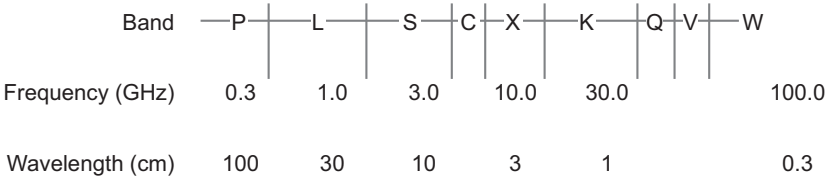


Figure 10.3: Microwave spectrum and band identification by letters.

Microwave polarizations The polarization of an electromagnetic wave is important in radar remote sensing. Depending on the orientation of the transmitted and received radar wave, polarization will result in different images (see Figure 2.2, which shows a vertically polarized EM wave). 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:

Polarization

- 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.

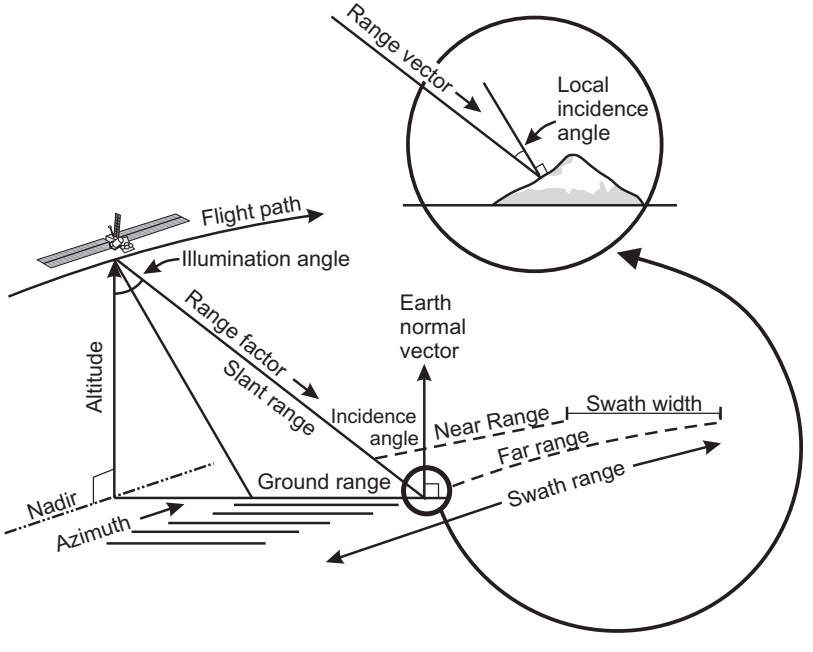


Figure 10.4: Radar remote sensing geometry.

10.2.3 Geometric properties of radar

The platform carrying the radar sensor moves along the orbit or flight path (see Figure 10.4). 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, *ie*, side-looking. The direction along-track is called *azimuth*, the direction perpendicular (across-track) is sometimes called *range*.

Azimuth

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 10.4). 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 10.4). 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 measured point in slant range is called *ground range* (Figure 10.4).

Ranges

Incidence angle

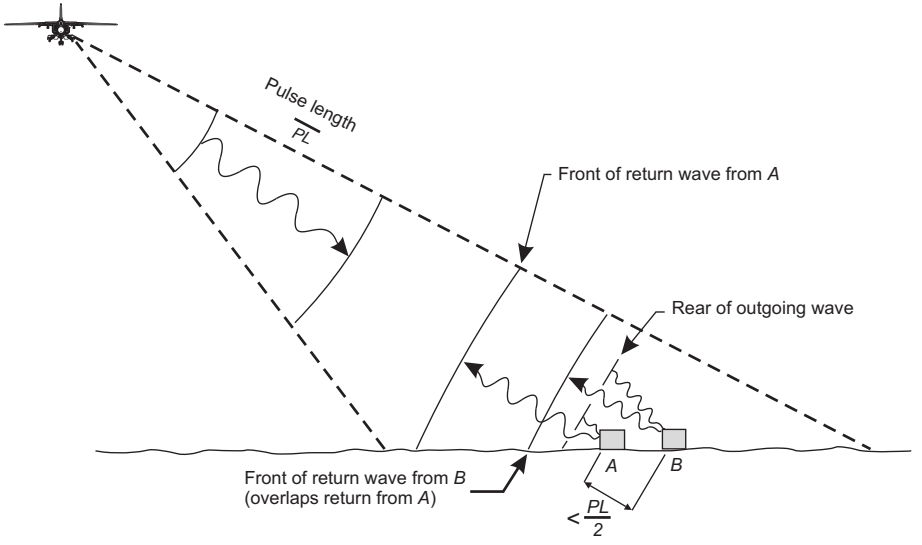


Figure 10.5: Illustration of the slant range resolution.

Spatial resolution

In radar remote sensing, the images are created from the backscattered portion of transmitted signals. Without further sophisticated processing, the spatial resolutions in slant range and azimuth direction are defined by the pulse length and the antenna beam width, respectively. This setup is called *real aperture radar* (RAR). Due to the different parameters that determine the spatial resolution in range and azimuth direction, 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 data to the same GSD in both directions.

RAR

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. Two objects can be resolved in range direction if they are separated by at least half a pulse length. In that case the return signals will not overlap. The slant range resolution is independent of the range (Figure 10.5).

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, *ie*, *aperture*; this means the longer the antenna, the narrower the beam and the higher the spatial resolution in azimuth direction.

Aperture

Synthetic Aperture Radar

RAR systems have their limitations in getting useful spatial resolutions of images because there is a physical limit to the length of the antenna, that can be carried on an aircraft or satellite. On the other hand, shortening the wavelength will reduce the penetrating capability of clouds. To improve the spatial resolution a large antenna is synthesized. The synthesization is achieved by taking advantage of the forward motion of the platform. Using all the backscattered signals in which a contribution of the same object is present, a very long antenna can be synthesized. This length is equal to the part of the orbit or the flight path in which the object is "visible". Most airborne and spaceborne radar systems use this type of radar. Systems using this approach are called *Synthetic Aperture Radar (SAR)*.

SAR

10.2.4 Data formats

SAR data are recorded in so-called raw format. It can be processed with a SAR processor into a number of derived products, such as intensity images, geocoded images and phase-containing data. The highest possible spatial resolution of the raw data is defined by the radar system characteristics.

Raw data

Raw data contain the backscatter of objects on the ground "seen" at different points in the sensor orbit. The received backscatter signals are sampled and separated into two components, together forming a complex number. The components contain information about the amplitude and the phase of the detected signal. The two components are stored in different layers. In this format, all backscatter information is still available in the elements of the data layers and:

- Each line consists of the sampled return signal of one pulse.
- An object is included in many lines (about 1000 for ERS).
- The position of an object in the different lines varies (different range).
- Each object has a unique Doppler history which is included in the data layers.

SLC data

The raw data are compressed based on the unique Doppler shift and range information for each pixel, which means that the many backscatters of a point are combined into one. The output of that compression is stored in one pixel which is still in complex format. Each pixel still contains information of the returned microwave. The phase and amplitude belonging to that pixel can be computed from the complex number. If all backscatter information of a point is used in the compression, then the output data is in Single Look Complex (SLC) format. The data still have their highest possible spatial resolution.

Multi-look data

In the case of multi-look processing, the total section of the orbit in which an object can be "seen" is divided into several parts. Each part provides a 'look' at the object. Using the average of these multiple looks, the final image is obtained which is still in complex format. Multi-look processing reduces the spatial resolution but it reduces unwanted effects (speckle) by averaging.

Intensity image

To get a visually interpretable image, the SLC or multi-look data need to be processed. The complex format is transformed into an 'intensity image'. In fact the norm (length) of the complex vector gives the "intensity of the pixel". The spatial resolution of the intensity image is related to the number of looks that are used in the compression step.

10.2.5 Distortions in radar images

Due to the side-looking viewing geometry, radar images suffer from serious geometric and radiometric distortions. In a radar image, you encounter variations in scale (caused by slant range to ground range conversion), 'foreshortening', 'layover' and 'shadows' (due to terrain elevation; Figure 10.6). Interference due to the coherence of the signal causes the 'speckle' effect.

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 (see Figure 10.4). This means that objects in near range are compressed with as compared with objects at far range. For proper interpretation, the image has to be corrected and transformed into ground range geometry.

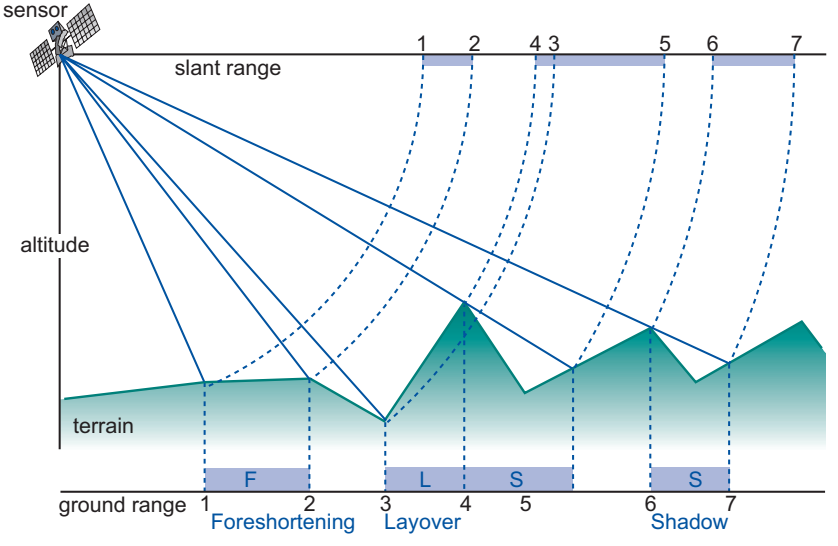


Figure 10.6: Geometric distortions in a radar image caused by varying terrain elevation.

Terrain-induced distortions

Similarly to optical sensors that can operate in an oblique manner (*eg*, SPOT) radar images are subject to relief displacements. In the case of radar, these distortions can be severe. There are three effects that are typical for radar: *foreshortening*, *layover* and *shadow* (Figure 10.6).

Foreshortening Radar measures distance in slant range. The slope area facing the radar is compressed in the image. The amount of shortening depends on the angle that the slope forms in relation to the incidence angle. 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, *ie*, 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.

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, *ie*, it is compressed into fewer 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. Slopes facing the radar appear (very) bright. Unfortunately this effect cannot be corrected for. This is why especially layover and shadow areas in a radar image 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 surface structure and terrain relief.

Speckle

Interference

A typical property of radar images is the so-called *speckle*. It appears as grainy "salt and pepper" effects in the image (Figure 10.7). Speckle is caused by the interference of backscattered signals coming from an area which is included in one pixel. 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 images difficult.

Speckle reduction

It is possible to reduce speckle by means of multi-look processing or spatial filtering. If you purchase an ERS SAR scene in "intensity (PRI)-format" you will receive a 3-look or 4-look image. 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 an image.

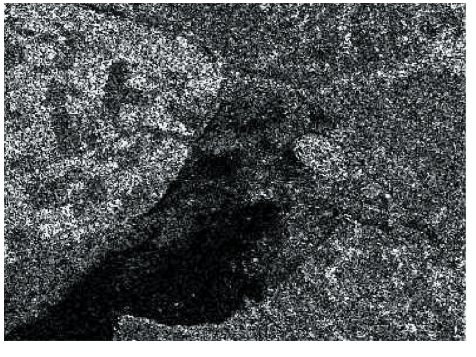
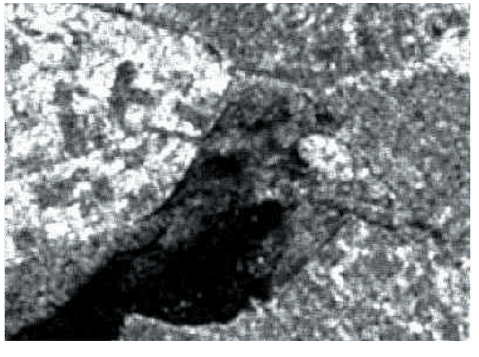
(a)   (b)

Figure 10.7: Original (a) and speckle filtered (b) radar image.

10.2.6 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.

Microwave signal and object interactions

For interpreters who are concerned with visual interpretation of radar images, the degree to which they can interpret an image depends upon whether they can identify typical/representative tones related to surface characteristics. 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, etc) and the characteristics of the illuminated object (roughness, shape, orientation, dielectric constant, etc).

Surface roughness is the terrain property that most strongly influences the strength of radar returns. It is determined by textural features comparable to the size of radar wavelength (typically between 5 and 40 cm), such as leafs and twigs of vegetation, and sand, gravel and cobble particles. A distinction should be made between surface roughness and terrain relief. Surface roughness occurs at the level of the radar wavelength (centimetres to decimetres). By terrain relief we mean the variation of elevation of the ground surface; relative to the resolution of radar images only elevation change in the order of metres is relevant. *Snell's law* states that the angle of reflection is equal and opposite to the angle of incidence. A smooth surface reflects the energy away from the antenna without returning a signal, thereby resulting in a black image. With an increase in surface roughness, the amount of energy reflected away is reduced, and there is an increase in the amount of signal returned to the antenna. This is known as the backscattered component. The greater the amount of energy returned, the brighter the signal is shown in the image. A radar image is, therefore, a record of the backscatter component and is related to surface roughness.

Surface roughness

Terrain relief

Complex dielectric constant Microwave reflectivity is a function of the complex dielectric constant. The complex dielectric constant is a measure of the electrical properties of surface materials. The *dielectric constant* of a medium consists of a part referred to as permittivity, and a part referred to as conductivity [31]. Both properties, permittivity and conductivity, are strongly dependent on the moisture or liquid water content of a medium. Material with a high dielectric constant has a strongly reflective surface. Therefore, the difference in the intensity of the radar return for two surfaces of equal roughness is an indication of the difference in their dielectric properties. In case of soils this could be due to differences in soil moisture content.

Surface Orientation Scattering is also related to the orientation of an object relative to the radar antenna. For example the roof of a building appears bright if it faces the antenna and dark if the incoming signal is reflected away from the antenna. Thus backscatter depends also on the local incidence angle.

Volume scattering is related to multiple scattering processes within a group of objects, such as the vegetation canopy of a wheat field or a forest. The cover may be all trees, as in a forested area, which may be of different species with variation in leaf form and size, or grasses and bushes with variations in form, stalk size, leaf and angle, fruiting and a variable soil surface. Some of the energy will be backscattered from the vegetated surface, but some, depending on the characteristics of radar system used and the object material, will penetrate the object and be backscattered from surfaces within the vegetation. Volume scattering is therefore dependent upon the heterogeneous nature of the object surface and the physical properties of the object such as leaf size, direction, density, height,

presence of lower vegetation, etc as well as the characteristics of the radar used, such as wavelength and related effective penetration depth [2].

Point objects are objects of limited size that give a very strong radar return. Usually the high backscatter is caused by the so called 'corner reflection'. An example is the dihedral corner reflector—a point object situation resulting from two flat surfaces intersecting at 90° and situated orthogonal to the radar incident beam. Common forms of dihedral configurations are man-made features, such as transmission towers, railroad tracks, or the smooth side of buildings on a smooth ground surface. Another type of point object is a trihedral corner reflector, which is formed by the intersection of three mutually perpendicular flat surfaces. Point objects of the corner reflector type are commonly used to identify known fixed points in an area in order to perform precise calibration measurements. Such objects can occur naturally and are best seen in urban areas where buildings can act as trihedral or dihedral corner reflectors. These objects give rise to intense bright spots on an image and are typical for urban areas. Point objects are examples of objects that are sometimes below the resolution of the radar system, but because they dominate the return from a cell they give a clearly visible point, and may even dominate the surrounding cells.

Corner reflection

10.2.7 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, etc. 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 instant (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. Radar data is also used for oil slick monitoring and environmental protection.

10.2.8 InSAR

Radar data provide a wealth of information that is not only based on a derived intensity image but also on other data that capture characteristics of the objects. One example is *SAR interferometry (InSAR)*, an advanced processing method that takes advantage of the phase information of the microwave. InSAR is a technique that enables the extraction of 3D information of the Earth's surface. It is based on the phase differences between corresponding pixels in two SAR images of the same scene but acquired at a slightly different position. The different path lengths from these positions to the target on the Earth's surface cause the differences in phase (Figure 10.8). SAR systems can detect the phase of the return signals very accurately.

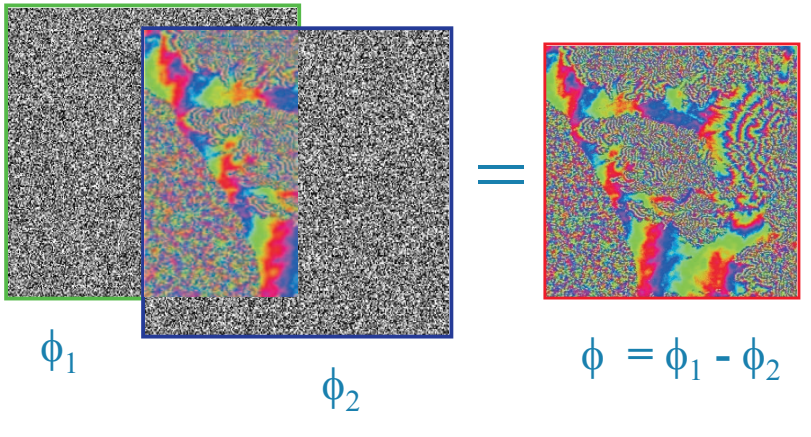


Figure 10.8: Phase differences forming an interferogram.

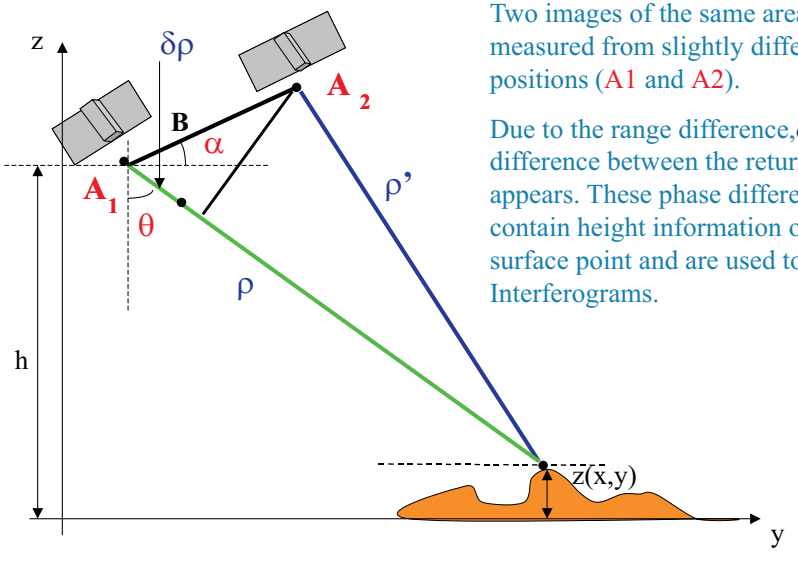
Data acquisition modes

Radar data for InSAR can be collected in two different modes:

- Single or simultaneous pass interferometry: in this mode, two images are simultaneously acquired from two antennas mounted on the same platform but separated by a distance known as baseline. This mode is mainly applied with aircraft systems but was also used for the Shuttle Radar Topographic Mission (SRTM), where receiving antennas were located at two ends of a mast of 60 metres as baseline length.

Baseline

- Repeat or dual pass interferometry: in this mode, two images of the same area are taken in different passes of the platform. The SAR data acquired from satellites such as ERS-1 and ERS-2, JERS-1 and RADARSAT may be used in this mode to produce SAR interferograms but also some aircraft systems are based on this mode.



Two images of the same area are measured from slightly different positions (A1 and A2).

Due to the range difference,δρ, a phase difference between the return signals appears. These phase differences contain height information of the surface point and are used to create the Interferograms.

Figure 10.9: Illustration of the InSAR geometry.

Concept

Phase difference
Interferogram

The phase information of two radar data sets (in SLC format) of the same region is used to create a DSM of the region (Figure 10.9). The corresponding elements of two SLC data sets are acquired at two slightly different antenna positions (A_1 and A_2). The connection between these positions forms the baseline. This baseline has a length (B) and an orientation (α) relative to the horizontal direction. The positions A_1 and A_2 are extracted from the platform orbits or flight lines. The difference in antenna position results in a difference in the range (ρ and ρ') from the target to those positions. This difference in range ($\delta\rho$) causes a phase difference (θ). This phase difference can be computed from the phases of corresponding elements in the SLC data sets, and are stored in the so-called *interferogram*. Finally, the terrain elevation is a function of the phase difference θ, the baseline B, some additional orbit parameters and it is represented in a chosen reference system.

Coherence

The basis of InSAR is phase comparison over many pixels. This means that the phase between scenes must be statistically similar. The coherence is a measure of the phase noise of the interferogram. It is estimated by window-based computation of the magnitude of the complex cross correlation coefficient of the SAR images. The interferometric coherence is defined as the absolute value of the normalized complex cross correlation between the two signals. The correlation will always be a number between 0 and 1. If corresponding pixels are similar then the correlation is high. If the pixels are not similar, *ie*, not correlated to a certain degree, then the phase will vary significantly and the coherence is low, meaning that the particular image part is de-correlated. Low coherence (*eg*, less than 0.1) indicates low phase quality and results in a noisy interferogram that causes problems in the DSM generation. The coherence decreases with increasing change in the random component of the backscattered fields between passes (temporal decorrelation) due to the physical change of the surface roughness structure reducing the signal correlation as do vegetation, water, shifting sand dunes, farm work (planting fields), etc. Geometric distortions caused by steep terrain slopes and orbit inaccuracies also decorrelate the images.

Decorrelation

10.2.9 Differential InSAR

Differential interferometry (DINSAR) using spaceborne sensors has become an established tool for the analysis of very small surface deformations. Its idea is to analyze the phase differences between SAR interferograms caused by surface displacements between the data acquisitions. Due to the short wavelength of SAR sensors, surface movements on the centimetre scale can easily be detected with an orbiting satellite from several hundred kilometres distance. However, atmospheric effects can also contribute to phase differences, which cannot be easily distinguished from surface displacements.

Concept

Three SAR images of the same area are acquired in three different passes to generate two interferograms. The arithmetic difference of the two interferograms is used to produce a differential interferogram. The elements of the differential interferogram, in combination with orbit information and sensor characteristics, is used to compute surface changes.

10.2.10 Application of DINSAR

One of the possibilities for interferometry using data from space is its possibility to derive global DEMs. Elevation mapping in areas such as the tropics was not feasible without radar's ability to penetrate cloud cover and to acquire images regardless of sunlight conditions. Utilizing ERS and RADARSAT images or SRTM data can lead to DEMs with absolute errors in elevation values of less than 10 metres. It must be noted, however, that computed elevation values from radar do not necessarily pertain to the ground surface. Horizontal resolution with ERS data of 20 metres and with RADARSAT Fine Beam mode 10 metres are possible. Georeferencing accuracy can be better than 20 metres. Moreover, the ability to create DSMs is beneficial for the generation of 3D images to assist in the identification of targets for military and environmental purposes. DTMs also provide critical information with reference to the geomorphic process. In this practice, interferometry is useful in detecting changes caused by alluvial fans, broad flood plain sedimentation patterns, sediment extraction, delta extensions and the formation and movement of large dune fields.

Elevation mapping

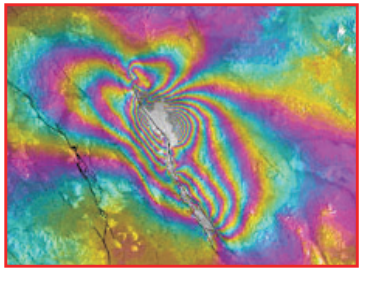


Figure 10.10: Surface deformation mapping with DINSAR.

Change detection

Change detection is an important field of study and is based on the practice of DINSAR. DINSAR allows for super-sensitive change detection with accurate measurements to even millimetres accuracy. This practice can be used for monitoring landslides and erosion. Moreover, DINSAR can be used to gather information on changes in areas where mining and water extraction have taken place (see Figure 10.10).

Co-seismic deformation

SAR can provide high-resolution images of earthquake-prone areas, high-resolution elevation data, and a high-resolution map of co-seismic deformation generated by an earthquake. Of these the last one is probably the most useful, primarily because it is unique. Other techniques are capable of generating images of the Earth's surface and elevation data, but no other technique provides high-spatial-resolution maps of earthquake deformation. Crust deformation is a direct manifestation of the processes that lead to earthquakes. Consequently, it is one of the most useful physical measurements we can make to improve estimates of earthquake potential. SAR interferometry can provide the required information.

Surface deformation by pressure

Interferograms were calculated to help study the activity of volcanoes through the creation of DEMs and mapping of land deformation. Researchers have used over 300 images of ERS-1 data to create many interferograms of Mount Etna (Italy). They were able to measure the increase in the size of the volcano (change detection and deformation) caused by the pressure of the magma in its interior. They were also able to follow the shrinking of the volcano once its activity had subsided as well as the changes in elevation of the surrounding area caused by the lava flows. This technique can be used to monitor awakening volcanoes to

prevent mass destruction, and for local and international relief planning.

Land use classification

Interferometry is useful in coherence-based land use classification. In this practice, the coherence of a repeat pass interferogram provides additional information contained in the radar backscatter. This information can be used as an input channel into land use classification. The use of coherence has proven successful for the separation of forests and open fields.

Ice monitoring

Interferometry is also useful in polar studies, including the measuring of flow velocities, tidal displacements and ice sheet monitoring. Researchers at the University of Alaska at Fairbanks were able to use interferometry to calculate the surface velocity field on the Bagley Icefield (Alaska) before and during the 1993-94 surge of the Bearing Glacier using ERS-1 data.

Waves and currents

In terms of ocean dynamics, interferometry can be used to study ocean wave and current dynamics, wave forecasting, ship routing, and placement and design of coastal and off shore installations.

Subsidence

Many major cities are located in areas undergoing subsidence as a result of withdrawal of ground water, oil or gas, and other minerals. Several metres of subsidence over several decades are not uncommon. Examples of cities with significant problems include Houston, Mexico City, Maracaibo, and Kathmandu. High rates of subsidence can have a major impact on flood control, utility distribution, and water supply.

Subsidence is also a result of natural processes such as limestone or marble dissolution that forms karst landscapes. In western Pennsylvania, an underground coal fire has burned for many years causing localized subsidence and threatening a much wider region with similar risks. Successive SAR images in urban areas over periods of several months may be able to detect subsidence directly. The surface structure of many parts of urban areas remains unchanged over several years, suggesting that interferometry over several years may be possible. Subsidence rates of several centimetres per year or more may be occurring in affected cities and should be detectable.

Satellite-based SAR interferometry has two important roles to play in polar studies. First, SAR interferometry can provide complete coverage elevation data. Second, repeat-pass interferometry can be used to measure ice flow and assess other changes. The cover image shows the first direct measurement of ice flow velocity from space without ground control.

10.2.11 Supply market

Spaceborne SAR interferometry holds great promise as a change detection tool in the fields of earthquake studies, volcano monitoring, land subsidence detection, and glacier and ice-stream flow studies. In a number of other fields, such as hydrology, geomorphology, and ecosystem studies, generating an accurate, globally consistent DSM with SAR interferometry is a major goal in itself.

The market for airborne interferometric elevation data generation is different from the one for airborne laser scanning in two aspects: (1) currently there is only one commercial company offering airborne InSAR services, and (2) elevation data are less accurate but much cheaper than those acquired by ALS.

10.2.12 SAR systems

The following systems mounted on airborne (Table 10.1) and spaceborne (Table 10.2) platforms did or do produce SAR data. However, not all of these systems generate data that can be processed into interferograms because of inaccuracies in orbit or flight pass data. This list, created in May 2004, is most probably not complete. For the latest situation refer to ITC's *Database of Satellites and Sensors*.

Instrument	Band/λ	Organization	Owner
Emisar	C,L band	Techn. Univ. of Denmark	Denmark
Pharus	C band	FEL-TNO	Netherlands
Star-31	X band	Intermap	Canada
Airsar/Topsar	P,L,C band	Nasa/JPL	USA
Carabas	3-15 cm	Chalmers University/FOI	Sweden
Geosar	X,P band	JPL and others	USA
WINSAR	4 bands	Metratec	USA

Table 10.1: Airborne SAR systems.

Instrument	Band	Remarks	Owner
Radarsat	C-band		Canada
ERS-1	C-band	Not operational anymore	ESA
ERS-2	C-band		ESA
Envisat	C-band		ESA
JERS -1	L-band	Not operational anymore	Japan
SRTM	C and X band	Space shuttle mission	NASA

Table 10.2: Spaceborne SAR systems.

10.2.13 Trends

The trend in airborne InSAR is towards multi-frequency and multi-polarization systems. The advantages of a long-wave-band (L or P) are that they can penetrate tree canopy and will probably result in a ground surface elevation map in dense forest. The use of combinations of short wavebands (X or C) with long wavebands enables biomass estimation. The use of multi-polarization InSAR enables the creation of optimized interferograms applying a weighted contribution of the different polarizations (HH, VH, HV, VV). The use of airborne SAR sensors for differential interferometry is also of great interest. The use of longer wavelengths with better coherence behaviour, such as L or P-band, offers the possibility of an analysis of long-term processes even in vegetated areas. The capability for monitoring of short-term processes is improved by the greater flexibility of airborne sensors. Particularly, the combination of space-borne interferometric SAR data with flexibly acquired airborne data is promising.

The future in spaceborne interferometry will be mainly in the direction of differential InSAR for several applications where change detection is important. In the coming years, two space borne systems will be launched, the Japanese PALSAR system on the ALOS satellite, and the German TerraSar providing radar data in higher spatial and spectral resolution modes.

10.3 Laser scanning

10.3.1 Basic principle

Laser scanning - in functional terms - can be defined as system that produces digital surface models. The system comprises an assemblage of various sensors, recording devices, and software. The core component is the laser instrument. The laser instrument measures distance, which is referred to as 'laser ranging'. When mounted on an aircraft, the laser rangefinder measures the distance to the terrain in very short time intervals. Combining a laser rangefinder with sensors that can measure the position and attitude of the aircraft (GPS & IMU) makes it possible to determine a model of the terrain surface in terms of a set of (X,Y,Z) coordinates, following the polar measuring principle (Figure 10.11).

Measuring by three sensors

We can define the coordinate system in such a way that Z refers to elevation. The digital surface model (DSM) thus becomes a digital elevation model (DEM), so we model the surface of interest by providing its elevation at many points with position coordinates (X,Y). Do the elevation values, which are produced by airborne laser scanning (ALS), refer to elevation of the 'bare ground' above a predefined datum? Not necessarily, since the 'raw DEM' gives us elevation of the surface the sensor "sees" (Figure 10.12). Post-processing is required to obtain a digital terrain relief model (DTM) from the DSM.

The key performance characteristics of ALS are: high ranging precision, yielding high resolution DSMs in near real time, and little or no dependence of flying on weather conditions, season, and light conditions. Typical applications of ALS are, therefore, forest surveys, surveying coastal areas and sand deserts, flood plain mapping, power line and pipeline mapping, monitoring open-pit mining,

Applications of ALS

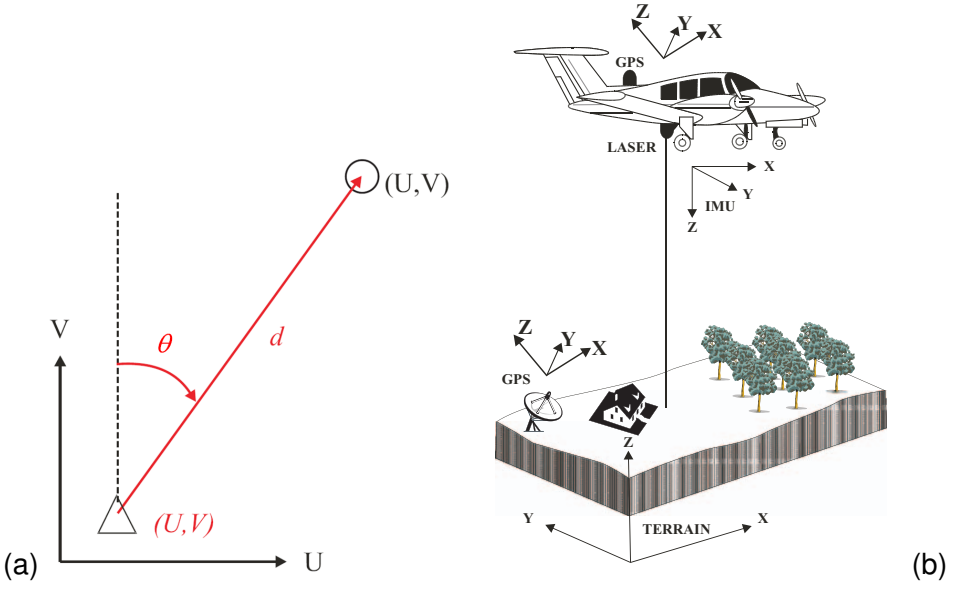


Figure 10.11: Polar measuring principle (a) and its application to ALS (b).

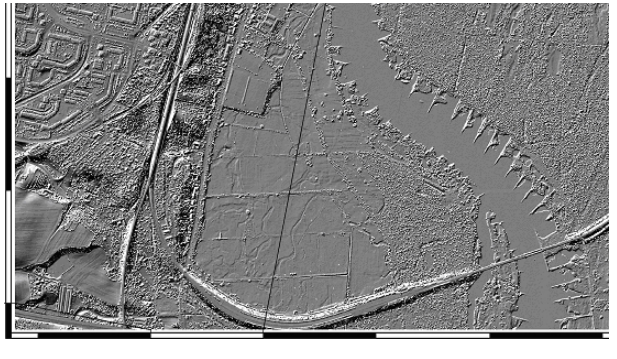


Figure 10.12: DSM of part of Frankfurt/Oder, Germany (1 m point spacing). Courtesy of TopoSys.

3D city modelling, etc.

10.3.2 ALS components and processes

Laser

LASER stands for Light Amplification by Stimulated Emission of Radiation. Einstein can be considered the father of laser, although he did not invent it. Roughly 85 years ago he postulated photons and stimulated emission; he won the Nobel Prize for related research on the photoelectric effect. In 1960, Theodore Maiman at Hughes Research Laboratories developed a device to amplify light, thus building the first laser (instrument). A laser emits a beam of monochromatic light, or radiation in the NIR range of the spectrum. The radiation is not really of a single wavelength, but it has a very narrow spectral band, smaller than 10 nm. Specific for a laser is also the very high intensity of the emitted radiation. Today lasers are used for many different purposes, among them for surgery. Lasers can damage cells (by boiling their water content), so they are a potential hazard to eye safety. Therefore, safety classes have been established for laser rangefinders, which must be observed for surveying applications.

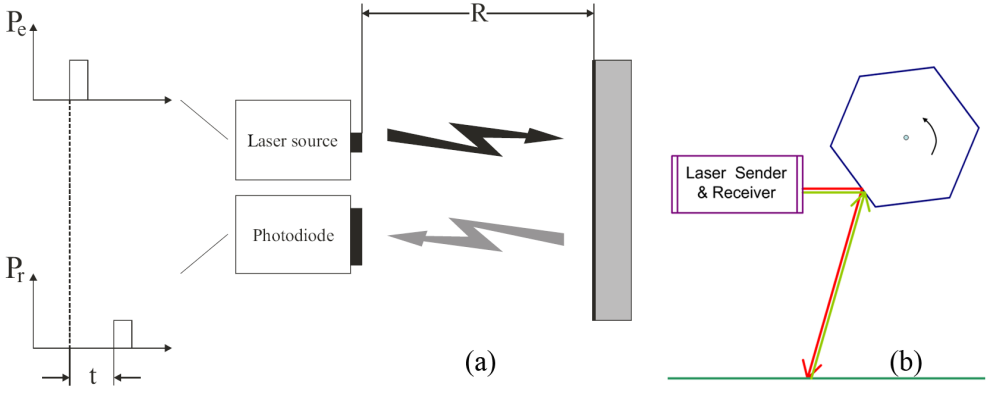


Figure 10.13: Concept of laser ranging and scanning [35].

Laser rangefinders and scanners come in various forms. Most airborne laser instruments are 'pulse lasers'. A pulse is a signal of very short duration and it travels as a beam. Airborne laser rangefinders for topographic applications emit NIR radiation. An emitted pulse is reflected on the ground and its return signal is sensed by a photodiode (Figure 10.13(a)). A time counter is started when a pulse is sent out and stopped on its return. The elapse time, measured with a resolution of 0.1 nanoseconds, can easily be converted to a distance as we know the speed of light, c:

Laser rangefinders

$$R = \frac{1}{2} c \cdot t. \tag{10.2}$$

Modern laser scanners send out pulses at a very high frequency (up to 250,000 pulses per second). Across-track scanning is in most cases achieved by a moving mirror, which deflects the laser beam (Figure 10.13(b)). The mirror can be of the oscillating, rotating, or nutating type. The Falcon system uses fiber optics to achieve scanning. Adding a scanning device to a ranging device has made surveying of a large area more efficient; a strip (swath of points) can be captured with a single flight line instead of a just a line of points as it was the case with the earlier versions of laser systems, the laser profilers.

Laser scanner

Simple laser rangefinders register one return pulse for every emitted pulse. Modern laser rangefinders for airborne applications record multiple echoes from the same pulse. Multiple return laser ranging is specifically relevant for flying terrain with vegetation, because it facilitates distinguishing vegetation echoes from ground echoes. The pulse may hit a leave at the top of a tree and part of it is reflected while part of it travels farther. It may then hit a branch and eventually

Multiple return ranging

even the ground (Figure 10.14). Many of the first return echos will be from the tree canopy, while the last returns are more likely to stem from the ground. Each return can be converted to (X,Y,Z) of the illuminated target point. To figure out whether the point is on the ground or somewhere on the vegetation is not trivial. Multiple return ranging does not give a direct answer but it helps finding it. An example of a first return and last return DSM is shown in Figure 10.15 A further development is 'full waveform sensors'. Full waveform scanners or altimeters (as on ICESat, see below) digitize the entire return signal of each emitted laser pulse instead of only detecting an echo if its intensity is above a certain threshold (see Figure 10.14). Full waveform laser rangefinders can provide information about surface roughness and more cues on vegetation cover.

Full waveform sensors

Along with measuring the range some laser instruments measure the amplitude of the reflected signal in order to obtain an image (often referred to as 'intensity image'). Imaging by a laser scanner is different from imaging by a radar instrument. While an image line of a microwave radar image stems from a single pulse, an image line of a laser intensity image stems from many pulses and is formed in the same way as for an across-track multispectral scanner. The benefit of "imaging lasers" is limited. The obtained images are monochromatic and are of lower quality than panchromatic images. A separate camera or multispectral scanner can produce much richer image content.

Imaging laser

ALS provides 3D coordinates of terrain points. To calculate accurate coordinates of terrain points we must accurately observe all needed elements. Measuring the distance from the aircraft to the terrain can be done very precisely by the laser rangefinder, within centimetres. We can accurately determine the position of the

GPS and IMU

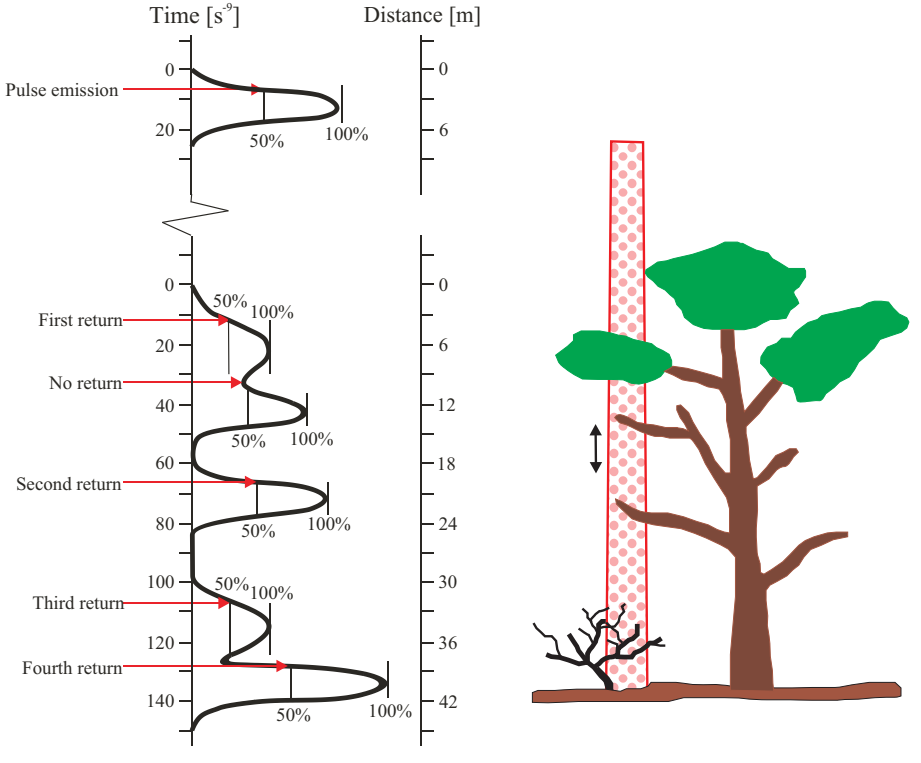


Figure 10.14: Multiple return laser ranging. Adapted from Mosaic Mapping Systems, Inc.

aircraft by differential GPS, using dual frequency receivers. To also accurately determine the attitude of the aircraft at the moment a distance is measured, we employ an IMU.

The total weight of the GPS-IMU-scanner equipment used to be about 150 kg

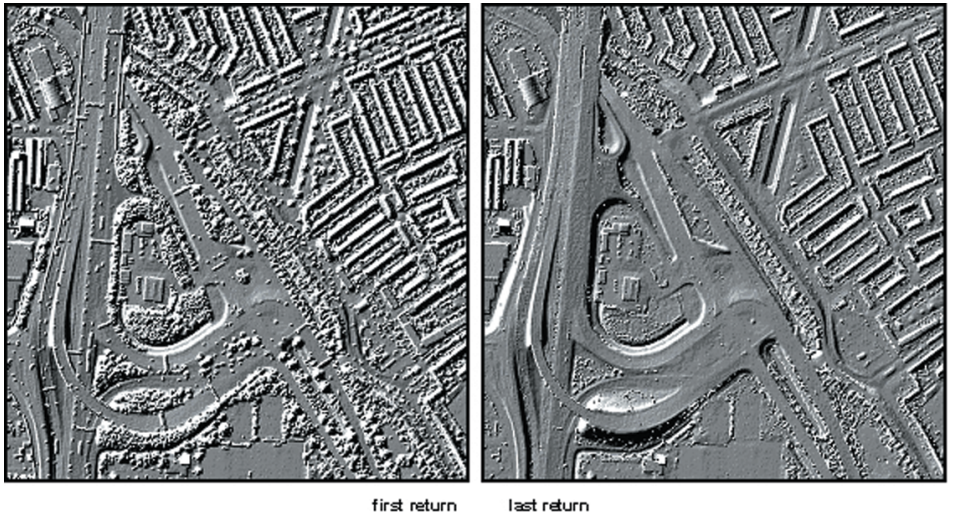


Figure 10.15: First and last return DSM of the same area. Courtesy of TopoSys.

four years ago. In 2008, a system like IGI's LiteMapper 5600 only weights 48 kg. Most ALS systems are used in a small to medium fixed wing aircraft. A typical platform would range from a Cessna 206/210 or equivalent, to a Piper Navajo or equivalent. Usually an aircraft that is equipped for an aerial photography mission can be used to fly current commercially available ALS systems. There are several ALS systems that can be mounted on helicopters, and several have been designed to be exclusive to this type of platform. Helicopters are better suited for very high-resolution surveys, because they can easily fly slowly. The minimum flying height is among other parameters dependent on the eye-safe distance of the laser. The major limiting factor of the maximum flying height is energy loss of the laser beam. 1000 m and less are frequent flying heights, yet there are systems that can be flown at 8000 m.

ALS platforms

Different from an aerial survey for a stereo coverage of photographs, where each terrain point should get recorded at least twice, in ALS a terrain point is only "collected" once in principle, even if we fly overlapping strips. This is of advantage for surveying urban areas and forests, but of disadvantage for error detection.

Co-registering the data

After the flight, the recordings from the laser instrument and the position and orientation system (*ie*, integrated GPS and IMU) are co-registered to the same time and then converted to (X, Y, Z) for each point that was hit by the laser beam. The resulting data set may still contain systematic errors and is often referred to as "raw data".

Extracting information

Further data processing has then to solve the problem of extracting information from the un-interpreted set of (X, Y, Z) coordinates. Typical tasks are "extracting buildings", modelling trees (*eg*, to compute timber volumes), and most prominently: to filter the DSM to obtain a DTM. Replacing the elevation value at non-ground points by an estimate of the elevation of the ground surface is also referred to as vegetation removal or short "devegging", a term being maintained from the early days when ALS was primarily used for forested areas (Figure 10.16).

Critical to getting the right data at the right time are proper system calibration, accurate flight planning and execution (including the GPS logistics), and adequate software.

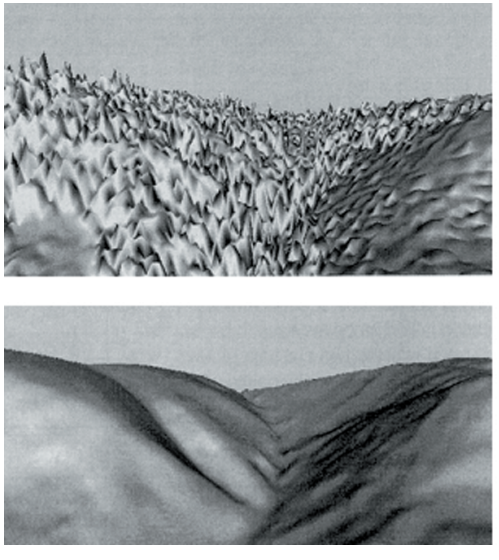


Figure 10.16: Devegging laser data: filtering a DSM to a DTM (from [14].

10.3.3 System characteristics

ALS produces a DSM directly comparable with what is obtained by image matching of aerial photographs/images. 'Image matching' is the core process of automatically generating a DSM from stereo images. Alternatively we can also use microwave radar to generate DSMs and eventually DTMs. The question is, why go for ALS?

There are several good reasons for using ALS for terrain modelling:

- A laser range finder measures distance by recoding the elapse time between emitting a pulse and receiving the reflected pulse from the terrain. Hence, the laser range finder is an active sensor and can be flown at day and night. The possibility of flying at night comes in handy, *eg*, for surveying a busy airport.
- Different from indirect distance measuring as done when using stereo images, laser ranging does not depend on surface/terrain texture.
- Laser ranging is less weather dependent than using passive optical sensors. A laser cannot penetrate clouds as microwave radar can, but it can be flown at low altitude, thus very often below the cloud ceiling.
- The laser beam is very narrow, with a beam divergence that can be less than 0.25 mrad; the area illuminated on the ground, therefore, can have a diameter smaller than 20 cm (depending on the laser type and flying

height). The idealization of "measuring points" is thus closely approximated. ALS can "see" objects that are much smaller than the footprint of the laser beam, therefore, we can use it for mapping power lines, etc.

- A laser beam cannot penetrate leaves, but it can pass through the tree canopy, unless very dense.
- A laser range finder can measure distances very precisely and very frequently; therefore, a DSM with a high density of points can be obtained and precise elevation values. The attainable elevation (vertical coordinate) accuracy by ALS can be in the order of 3 cm for well defined target surfaces.
- The multiple return recording facility offers "feature extraction", especially for forest applications and urban mapping (building extraction), both subjects of vivid research still.
- Moreover, the entire data collection process is digital and can be automated to a high degree, thus facilitating fast production.
- Another important advantage is that ALS does not need ground control other than a calibration site, which can usually be created near the airfield.

There are two major advantages of laser ranging compared to microwave radar: high energy pulses can be generated in short intervals and highly directional beams can be emitted. The latter is possible because of the short wavelength of lasers (10,000 to 1,000,000 times shorter than microwave). The consequence is much higher ranging accuracy.

Note that the term 'radar' is often used for short for microwave radar; however, in the literature you may also come across the term "laser radar", which is a synonym for laser ranging. A more frequently used synonym for laser ranging is LIDAR, although there are also lidar instruments that do not measure distance but the velocity of a target ('Doppler lidars').

10.3.4 Variants of Laser Scanning

Laser bathymetry

The first airborne laser ranging experiments were conducted in North America in the 1970ies aiming at bathymetric applications. There are currently a half dozen airborne laser bathymetry systems in operation. These are heavy and very expensive systems commonly employing lasers of two different wavelengths, near-infrared or green. Near infrared light has good atmospheric penetration but does not penetrate water or the ground. The near infrared beam is reflected by a water surface, while the green beam penetrates (clear) water and is reflected from the bottom of the water body. Measuring the time difference between the returns of the co-aligned laser pulses allows determining the water depth (shallow water, not deeper than 80 m). For precise topographic surveys (using near infrared lasers), progress was first required in satellite positioning (GPS in particular), which allowed the development of laser profiling in the late 1980s.

Terrestrial laser scanners

The idea of creating surface models and true 3D models by laser ranging can also be applied to problems where an aircraft does not offer the right perspective, and where ground-based cameras would fail to do a good job (*eg*, because the objects/scene to be surveyed have/has little texture). Ground-based laser scanners, also called terrestrial laser scanners (TLSs), combine the concepts of tacheometry (polar measuring, *eg*, by Total Station; explicit distance) and photogrammetry (bundle of rays, implicit distance). TLSs (*eg*, Leica Cyrax 2500, Riegl 3D-Imaging Sensor LMS-Z210) do not require a prism/reflector at a target point and can yield a very high density of points in a very short time. Such instruments may have a scanning range of 1 to 200 m and record 1000 points per second with an accuracy of 2 to 6 mm.

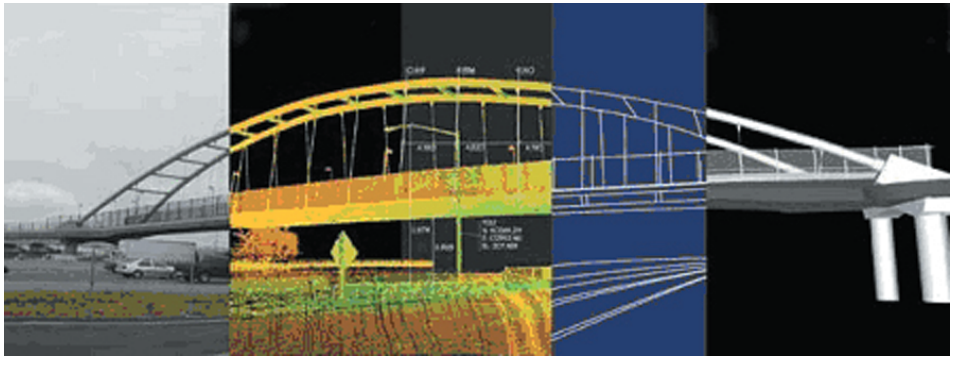


Figure 10.17: Images related to various stages of 3D modelling by a TLS (Leica).

TLSs are particularly suited for surveying in dangerous or inaccessible environments and for high precision work. All kinds of civil engineering, architectural, and archaeological surveys are the prime application areas. Figure 10.17 illustrates a typical TLS process chain. The TLS scans an object, usually from different positions to enable true 3D modelling of the object of interest (*eg*, a bridge, a building, a statue). For each viewing position of the TLS we obtain a point cloud, which can be shown as a 2D picture by colour coding the ranges to the object's surface. We can also co-register all point clouds into one common 3D coordinate system and then fit simple geometric shapes (cylinders, spheres, cubes, etc) to the (X,Y,Z) data. This way we can create CAD models and use computer aided design software, *eg*, for intelligible visualizations. An even more recent application area of ground-based laser scanning is mobile mapping and automatic vehicle navigation.

There is also satellite laser ranging, either used to measure the distance from

a ground station to a satellite with high precision, or to observe targets on the ground from spy satellites. NASA had several laser altimeters on board ICESat with a particular interest in ice and clouds. Only one of the full waveform lasers still operates in 2008, producing data which are also used for determining land elevation and for vegetation studies.

Satellite laser ranging

10.3.5 Supply Market

The acceptance of ALS data has grown. While in the early days of ALS, acquiring ALS data was a matter of contracting a private company, you can now buy ALS hardware and software and try to make it all work yourself. A potential client of ALS products is likely to have several options (in Europe and Northern America): (a) purchasing a total service by a private company for getting a tailor made DSM and products derived from it, (b) purchasing a value-added product from a reseller (*eg*, derived from a national DSM), (c) purchase a part of a national DSM and process it further in-house, and (d) buying a system and operating it.

The first country to establish a DSM for the entire territory using ALS was The Netherlands. The product (AHN) became available for any part of the country by 2004–with a density of at least 1 point per per 16 m^2 and 1 point per 32 m^2 in forest areas. Currently AHN2 is generated; the new survey will yield elevation data with a density of 10 points per m^2. Several other states in Europe have followed the Dutch example and have created a countrywide high density DSM based on ALS.

With the increasing availability of systems and services, researchers look into possibilities of refined processing (*eg*, further reduction of systematic errors and elimination of blunders) and automated post-processing. The latter refers to attempts to derive a DTM from the DSM, and to classify and delineate terrain features, in particular buildings. Another research topic concerns the derivation of terrain breaklines. Meanwhile, the manufactures will continue to aim for higher

measuring speeds and higher laser power, so that higher density DSMs can be produced from higher altitudes. Another trend is toward multimodal 3D systems, in particular integrating a laser scanner with a line camera (*eg*, the ALS80 + ADS 60 of Leica or the FALCON III system of TopoSys, where the laser scanner and an RGB–NIR scanner are assembled into a singe sensor rack), or with a frame camera (*eg*, Optech ALTM Gemini 167).

Summary

In this chapter, the principles of imaging radar, interferometric SAR, laser scanning and their respective applications have been introduced. The microwave interactions with the surface have been explained to illustrate how radar images are generated, and how that can be 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 interpretability. Radar data have many potential applications in the fields of geology, oceanography, hydrology, environmental monitoring, land use and land cover mapping and change detection. The concept and applications of InSAR were also explained, including how differential InSAR allows the detection of surface deformation.

In the second part the principle of laser scanning and its historical development have been outlined. The principal product is a digital surface model. Capabilities of airborne laser scanning and operational aspects have been introduced, and current trends reviewed. Prime applications of laser scanning data are forest surveys, surveying of coastal areas and sand deserts, flood plain mapping, power line and pipeline mapping, and 3D city modelling.

Additional reading

[7], [21].

Questions

The following questions can help to study Chapter 10.

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.
6. What are the components of an airborne laser scanning system, and what are the operational aspects to consider in planning and executing an ALS mission?
7. What are the key performance characteristics of ALS, and what are the major differences with airborne microwave radar?
8. What makes ALS especially suited for the mentioned applications?

Chapter 11

Image restoration and atmospheric corrections

11.1 Introduction

Chapter 5 introduced the notion of radiometric correction. Now we will first have a closer look at 'cosmetic corrections', which belong to the category image restoration, and then then treat in more detail relative and absolute atmospheric correction. Relative atmospheric correction is based on ground reflectance properties, whereas absolute atmospheric correction is based on atmospheric process information. Before explaining how to correct, the Section 11.2 will review the imaging process and the occurring disturbances.

The radiance values of reflected polychromatic solar radiation and/or the emitted thermal and microwave radiances from a certain target area on the Earth's surface are for researchers the most valuable information obtainable from a remote sensor. In the absence of an atmosphere, the radiance for any wavelength at the ground would be the same as the radiance at the sensor. No atmosphere would make RS easier, but living impossible. So we have to figure out how we can convert remotely detected radiances to radiances at the ground.

11.2 From satellite to ground radiances

The presence of a heterogeneous, dense and layered terrestrial atmosphere composed of water vapour, aerosols and gases disturbs the signal reaching the sensor in many ways. Therefore, methods of atmospheric corrections (AC) are needed to "clean" the images from these disturbances, in order to allow the retrieval of pure ground radiances from the target. The physics behind the AC techniques in the visible and in the thermal range are essentially the same, meaning that the same AC procedures applicable in one also apply to the other. However, there are a number of reasons and facts that allow a distinction between techniques applicable to visible and thermal data:

- Incident and reflected solar radiation and terrestrial thermal emission fall into very different parts of the spectrum.
- Solar emission and reflection depend on the position of the sun and the satellite at the moment of image acquisition. Thermal emission is theoretically less dependent on this geometry.
- Solar energy travels twice through the atmosphere before it reaches the sensor (Top of the Atmosphere (TOA) - ground - sensor), whereas ground thermal emissions only pass the atmosphere once (ground - sensor; see Figure 2.8).
- Solar reflection at the Earth's surface depends on material reflectance (ρ). Thermal emissions from the Earth depends on the emissivity of the surface materials (ϵ). Since solar reflection and earth thermal emission occur

in different wavelengths, the behaviour of one is not an indication of the other.

- The processes of atmospheric attenuation, *ie*, scattering and absorption, are both wavelength dependent and affect the two sectors of the spectrum differently.
- Because of the previous statement, AC techniques are applied at monochromatic level (individual wavelengths). It means that attenuation of energy is calculated at every individual wavelength and then integrated in the spectrum of the sensor by mathematical integration.
- Atmospheric components affect different areas of the spectrum in different ways, meaning that some components can be neglected when dealing with data of the thermal or the visible part of the spectrum.

A classification of different AC methods allows us to assess what kind of effort is needed to correct raw data for an application at hand. Some RS applications do not require AC procedures at all, except for some cosmetics, while some call for rigorous and complex procedures. Many applications can do with intermediate solutions.

In general, applications where the actual radiance at ground level is not needed do not require atmospheric correction. Some cosmetic and/or image enhancements procedures may suffice. Among them are mapping applications, where visual interpretation is important and image geometry, but not chemical properties of surface material.

Applications that require the quantification of radiation at ground level must include rigorous atmospheric correction procedures. Quantification of evapotranspiration or CO_2 sequestration, or surface temperature and reflectivity mapping are examples.

In between there are applications concerned with the evolution of certain parameters or land properties over time, rather than their absolute quantification. In those cases knowledge of the relative trend may suffice. These procedures apply mainly when the mapping parameters do not really have a meaningful physical value, simply because they were designed primarily for multitemporal relative comparison. Index evolution and correlation procedures, where radiances are associated with the evolution of certain parameter (*ie*, turbidity), are examples of this category. Be aware that some indexes such as NDVI typically require some absolute atmospheric correction.

The 'effort' required is synonymous with the amount of information required to describe the components of the atmosphere at different altitudes (atmospheric profiling) at the moment and position the image is taken, and less so with sophistication of the AC procedure itself. State-of-the-art atmospheric models allow the "cleaning" of any cloudless image regardless the sensor type, as long as atmospheric profile data are available. Unfortunately, such detailed atmospheric information can only be obtained through atmospheric sounding procedures, using of a series of instruments able to sample the atmosphere at fixed intervals while transported vertically by a balloon, or using satellite based sounding sensors. This kind of profiling is carried out daily at some atmospheric centres at fixed times, regardless of the satellite overpass time. However, the atmosphere

is dynamic. Atmospheric processes and composition change rapidly, mainly at low altitude (water vapour and aerosols), meaning that sounding made somewhere close to the target and near the time of the satellite overpass might not be enough to ensure an adequate atmospheric description. As a rule of thumb regarding AC techniques, first consider the objectives of the project, then identify the appropriate AC procedure, and finally establish the effort, *ie*, the required information to execute the chosen correction procedure.

11.3 Cosmetic corrections

The objective of what is called here cosmetics is to correct visible errors and noise in the raw data. No atmospheric model of any kind is involved at all in these correction processes; instead, corrections are achieved using especially designed filters and contrast stretching and enhancement procedures. These corrections are mostly executed (if required) at the satellite data receiving stations or image pre-processing centres, before reaching the final user. All applications require this form of correction.

Typical problems requiring cosmetic corrections are:

- periodic line dropouts;
- line striping;
- random noise or spike corrections.

These effects can be identified visually and automatically, and are here illustrated on a Landsat-7 ETM+ image of Enschede (Figure 11.1).

(a)

(b)

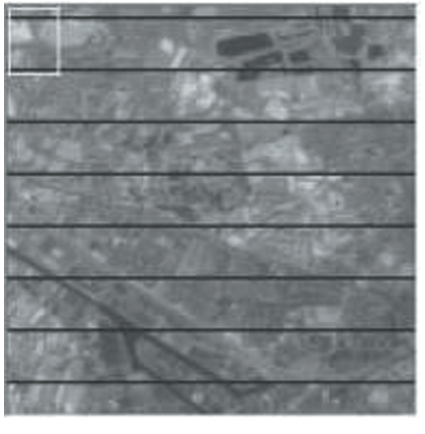
96	87	83	94	114	109	107	104	94	83	81	96	90	110	109
89	89	96	94	98	108	111	110	94	80	88	79	73	107	121
104	108	93	92	97	94	86	81	75	83	91	75	75	110	131
114	97	93	86	96	100	92	88	88	92	93	92	105	110	106
131	143	107	101	122	127	88	86	87	90	90	92	124	122	116
122	155	147	115	145	155	101	91	92	95	102	119	131	128	132
112	127	148	151	159	155	134	91	92	91	106	126	121	112	126
137	152	162	164	161	157	150	98	86	96	118	127	104	121	122
138	149	155	159	168	162	154	123	87	97	126	123	108	125	121
131	137	139	141	148	158	159	137	96	92	122	117	114	121	121
108	132	136	134	135	124	115	106	96	88	90	96	113	120	116
88	101	113	100	94	93	90	84	89	79	82	91	88	102	114
78	68	81	88	73	76	76	78	79	72	76	83	86	86	91
70	85	111	118	100	89	86	85	84	81	73	68	73	79	85
87	114	128	127	119	96	87	86	80	74	69	72	76	80	81
106	126	125	127	108	88	92	87	80	80	71	75	77	84	68
106	116	123	119	89	95	99	95	82	76	79	79	79	71	69
99	116	123	119	89	95	99	93	85	74	69	66	70	71	78
88	95	99	96	112	76	104	93	85	74	69	66	70	71	85
91	91	85	87	80	97	115	105	94	87	78	70	73	65	85

Figure 11.1: Original Landsat ETM image of Enschede and and its surroundings (a), and corresponding DNs of the the indicated subset (b).

11.3.1 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-7 ETM, for example, has 16 detectors for each of its channels, except the thermal channel. 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 in the image (Figure 11.2).

(a)



(b)

96	87	83	94	114	109	107	104	94	83	81	96	90	110	109
89	89	96	94	98	108	111	110	94	80	88	79	73	107	121
104	108	93	92	97	94	86	81	75	83	91	75	75	110	131
0	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
131	143	107	101	122	127	88	86	87	90	90	92	124	122	116
122	155	147	115	145	155	101	91	92	95	102	119	131	128	132
112	127	148	151	159	155	134	91	92	91	106	126	121	112	126
137	152	162	164	161	157	150	98	86	96	118	127	104	121	122
138	149	155	159	168	162	154	123	87	97	126	123	108	125	121
131	137	139	141	148	158	159	137	96	92	122	117	114	121	121
108	132	136	134	135	124	115	106	96	88	90	96	113	120	116
88	101	113	100	94	93	90	84	89	79	82	91	88	102	114
78	68	81	88	73	76	76	78	79	72	76	83	86	86	91
70	85	111	118	100	89	86	85	84	81	73	68	73	79	85
87	114	128	127	119	96	87	86	80	74	69	72	76	80	81
106	126	125	127	108	88	92	87	80	80	71	75	77	84	68
106	116	123	119	89	95	99	95	82	76	79	79	79	71	69
99	116	123	119	89	95	99	93	85	74	69	66	70	71	78
88	95	99	96	112	76	104	93	85	74	69	66	70	71	85
0	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Figure 11.2: The image with periodic line dropouts (a) and the DNs (b). All erroneous DNs in these examples are shown in bold.

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 his-

togram 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 (Figure 11.3). This restoration method is equally effective for random line dropouts that do not follow a systematic pattern.

(a)

(b)

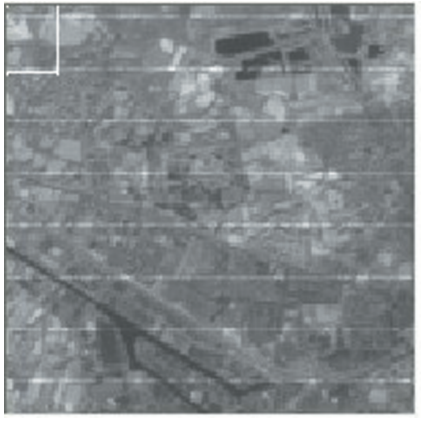
96	87	83	94	114	109	107	104	94	83	81	96	90	110	109
89	89	96	94	98	108	111	110	94	80	88	79	73	107	121
104	108	93	92	97	94	86	81	75	83	91	75	75	110	131
118	**126**	**100**	**97**	**110**	**111**	**87**	**84**	**81**	**87**	**91**	**84**	**100**	**116**	**124**
131	143	107	101	122	127	88	86	87	90	90	92	124	122	116
122	155	147	115	145	155	101	91	92	95	102	119	131	128	132
112	127	148	151	159	155	134	91	92	91	106	126	121	112	126
137	152	162	164	161	157	150	98	86	96	118	127	104	121	122
138	149	155	159	168	162	154	123	87	97	126	123	108	125	121
131	137	139	141	148	158	159	137	96	92	122	117	114	121	121
108	132	136	134	135	124	115	106	96	88	90	96	113	120	116
88	101	113	100	94	93	90	84	89	79	82	91	88	102	114
78	68	81	88	73	76	76	78	79	72	76	83	86	86	91
70	85	111	118	100	89	86	85	84	81	73	68	73	79	85
87	114	128	127	119	96	87	86	80	74	69	72	76	80	81
106	126	125	127	108	88	92	87	80	80	71	75	77	84	68
106	116	123	119	89	95	99	95	82	76	79	79	79	71	69
99	116	123	119	89	95	99	93	85	74	69	66	70	71	78
88	95	99	96	112	76	104	93	85	74	69	66	70	71	85
90	**93**	**93**	**90**	**102**	**95**	**110**	**101**	**87**	**81**	**77**	**74**	**74**	**81**	**95**

Figure 11.3: The image after correction for line dropouts (a) and the DNs (b).

11.3.2 Line striping

Line striping is far more common than line dropouts. 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 (Figure 11.4). It is important to understand that valid data are present in the defective lines, but these must be corrected to match the other lines of the image.

(a)



(b)

96	87	83	94	114	109	107	104	94	83	81	96	90	110	109
89	89	96	94	98	108	111	110	94	80	88	79	73	107	121
104	108	93	92	97	94	86	81	75	83	91	75	75	110	131
150	**127**	**121**	**113**	**129**	**130**	**121**	**116**	**115**	**120**	**127**	**124**	**143**	**148**	**142**
131	143	107	101	122	127	88	86	87	90	90	92	124	122	116
122	155	147	115	145	155	101	91	92	95	102	119	131	128	132
112	127	148	151	159	155	134	91	92	91	106	126	121	112	126
137	152	162	164	161	157	150	98	86	96	118	127	104	121	122
138	149	155	159	168	162	154	123	87	97	126	123	108	125	121
131	137	139	141	148	158	159	137	96	92	122	117	114	121	121
108	132	136	134	135	124	115	106	96	88	90	96	113	120	116
88	101	113	100	94	93	90	84	89	79	82	91	88	102	114
78	68	81	88	73	76	76	78	79	72	76	83	86	86	91
70	85	111	118	100	89	86	85	84	81	73	68	73	79	85
87	114	128	127	119	96	87	86	80	74	69	72	76	80	81
106	126	125	127	108	88	92	87	80	80	71	75	77	84	68
106	116	123	119	89	95	99	95	82	76	79	79	79	71	69
99	116	123	119	89	95	99	93	85	74	69	66	70	71	78
88	95	99	96	112	76	104	93	85	74	69	66	70	71	85
125	**121**	**111**	**119**	**107**	**132**	**160**	**139**	**128**	**121**	**104**	**93**	**99**	**90**	**113**

Figure 11.4: The image with line striping (a) and the DNs (b). Note that the destriped image would look 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 DNs are computed and assigned. This yields a destriped image in which all DN-values conform to the reference level and scale.

11.3.3 Random noise or spike noise

The periodic line dropouts and striping are forms of non-random noise that may be recognized 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 11.5). In the image these pixels produce bright and dark spots that interfere with information extraction procedures.

A spike noise can be detected by comparing neighbouring pixel values. If neighbouring pixel values differ by more than a specific threshold, the central pixel is considered spike noise and the DN is replaced by an interpolated DN.

(a)

(b)

96	87	83	94	114	109	107	104	94	83	81	96	90	110	109
89	89	96	94	98	108	111	110	94	80	88	79	73	107	121
104	108	93	92	97	94	86	81	75	**255**	91	75	75	110	131
114	97	93	86	96	100	92	88	88	92	93	92	105	110	106
131	143	107	101	122	127	88	86	87	90	90	92	124	122	116
122	155	147	115	145	155	101	91	92	95	102	119	131	128	132
112	127	148	151	159	155	134	91	92	91	106	126	121	112	126
137	152	162	**0**	161	157	150	98	86	96	118	127	104	121	122
138	149	155	159	168	162	154	123	87	97	126	123	108	125	121
131	137	139	141	148	158	159	137	96	92	122	117	114	121	121
108	132	136	134	135	124	115	106	96	88	90	96	113	120	116
88	101	113	100	94	93	90	84	89	79	82	91	88	102	114
78	68	81	88	73	76	76	78	79	72	76	83	86	86	91
70	85	111	118	100	89	86	85	84	81	73	68	73	79	85
87	114	128	127	119	96	87	86	80	74	69	72	76	80	81
106	126	125	127	108	88	92	87	**255**	80	71	75	77	84	68
106	116	123	119	89	95	99	95	82	76	79	79	79	71	69
99	116	123	119	89	95	99	93	85	74	69	66	70	71	78
88	95	99	96	112	76	104	93	85	74	69	66	70	71	85
91	91	85	87	80	97	115	105	94	87	78	70	73	65	85

Figure 11.5: The image with spike errors (a) and the DNs (b).

11.4 Atmospheric corrections

11.4.1 Relative AC methods based on ground reflectance

Relative AC methods avoid the evaluation of atmospheric components of any kind. They rely on the fact that for one sensor channel the relation between the radiances at TOA and at ground level follows a linear trend for the variety of earth features present in the image. This linear relation is in fact an approximation to reality, but precise enough to solve practical applications where there are other more important sources of errors. The methods are:

Two reflectance measurements: the output of this method is an absolute atmospherically corrected image, so it can be used on an individual basis, for multitemporal comparison or parameter evolution, and for flux quantification. Absolute means that the image output has physical units, and the calculated ground radiances are compatible with the actual atmospheric constituents. The application of this method requires the use of a portable radiometer able to measure in the same wavelength range as the band of the image to be corrected. If many bands are to be corrected, then the radiometer should have filters allowing the measurements in all these individual bands separately.

The procedure is straightforward. Prior to the satellite pass, some bright and dark sample areas in the image, preferably having the size of more than 3 pixels, are selected. It is not necessary that these targets are 'reflective invariant', although it is preferable if they are of uniform land cover. Reflective invariant areas are those that retain their reflective properties over time, *eg*, deep reservoir lakes. Another fundamental condition is that these pixels are univocally recognizable in the image. During the satellite pass, the reflectance of these targets is

measured in all bands in the field with the radiometer. After image acquisition the bands are individually calibrated, and the TOA radiance is read for these sample pixels. Plotting the ground radiance as a function of the TOA radiance for the target pixels, a linear equation expressing the AC linearity is found. This equation is then applied to the whole original TOA image to obtain the corrected image.

Two reference surfaces: The output of this method is an image that matches a reflectance that is compatible with the atmosphere of a similar image taken on a previous date. No absolute values of radiances are obtained in any of the two images, only allowing comparative results. This method works on an individual band/channel basis, and is valid to establish a uniform comparison basis to study, for example, the evolution of non-flux related parameters like indexes, or when certain convenient land properties can be derived directly or indirectly from the normalized radiance values in a band. The method relies on the existence of at least one dark and one bright invariant area. Normally a sizable area should avoid mixed pixels (mixed land cover). As rule of thumb it should be a minimum of 2 or 3 times larger than the image spatial resolution. Reflective invariant areas are considered to retain their reflective properties over time. Deep reservoir lakes, sandy beaches or desserts, open quarries, large salt deposits, big asphalted areas, are examples. It is supposed that for these pixels the reflectance should always be the same, since their composition keeps the reflective properties with time. If a difference in reflectance occurs at the reflective invariant area in the two date images, it can only be attributed to the different state of the atmosphere on these dates. The atmospheric composition is unknown in the two images, but their influence is measurable by calculating the difference in radiance between the two dates on the reflective invariant areas.

The procedure defines one image as master and the other as slave. The reflectance of one invariant pixel at the master image on date 1 is ρ_M, and the reflectance of the same pixel in the slave image on other date 2 is ρ_S. If the atmosphere is the same, then $\rho_M = \rho_S$. If not, the difference in reflectance produced by the distinctive atmospheric conditions can be measured as $\Delta\rho = \rho_M - \rho_S$. $\Delta\rho_i$ can be calculated for all i invariant reflective pixels found in the image, although two extremes (bright and dark), $\Delta\rho_b$ and $\Delta\rho_d$, will suffice. A linear relation, $\Delta\rho = a \cdot \rho_s + b$ is built out of these extreme invariant pixels. This equation $\Delta\rho = f(\rho_s)$ can then be used to correct the slave image to ensure that the invariant pixels match the reflectance of the master image. This is done in a new slave image, $\rho'_S = \rho_S + \Delta\rho = \rho_S + a \cdot \rho_S + b$. At the end ρ'_S is then an image showing the reflection of the second date image having the atmosphere of the first. In other words, the atmosphere of the master image was artificially imposed to the slave image, so that the evolution of reflectance can now be compared in both images, since they have the same atmosphere.

11.4.2 Absolute AC methods based on atmospheric processes

These methods require a description of the components in the atmospheric profile. The output of these methods is an image that matches the reflectance of the GRCs with a maximum estimated error of 10 %, if atmospheric profiling is adequate enough. This image can be used for flux quantifications, parameter evolution assessments, etc, as mentioned above. The advantage of these methods is that ground reflectance can be evaluated under any atmospheric condition, altitude and relative geometry between sun and satellite. The disadvantage is that the atmospheric profiling required for these methods is rarely available. Regarding this inconvenience, a sub-classification of absolute AC methods could be based on the accuracy of the method related to the effort in obtaining the required data.

	LOWTRAN/ MODTRAN	6S
Numerical approximation method(s)	Two-stream, including atmospheric refraction; discrete ordinates also in MODTRAN-3	Successive orders of scattering
Spectral resolution	20 cm^{-1} (LOWTRAN); 2 cm^{-1} (MODTRAN)	10 cm^{-1}, shortwave only
Clouds	Eight cloud models; user-specified optical properties	No clouds
Aerosols	Four optical models	Six optical models plus user-defined
Gas absorption [1]	Principle and trace gases	Principle and trace gases
Atmospheric profiles	Standard and user-specified	Standard and user-specified
Surface characteristics	Lambertian, no built-in models	Lambertian spectral albedo models built-in; bidirectionally reflecting surface possible
Primary output parameter	Radiance	Radiance/reflectance
User interface	Formatted input file	Input file

Table 11.1: Characteristics of LOWTRAN/MODTRAN and 6S RTMs. http: stratus.ssec.wisc.edu

Radiative transfer models

Radiative transfer models (RTMs) can be used for computing either radiances (intensities) or irradiances (fluxes) for a wide variety of atmospheric and surface conditions. They require a full descriptions of the atmospheric components at fixed altitudes throughout the atmosphere. LOWTRAN ([12]), MODTRAN ([1]), Code 5S ([28]) and 6S ([33]) are all reference RTMs. MODTRAN is becoming the standard for research studies in both the thermal and the visible spectrums. LOWTRAN was developed first, but was later superseded by MODTRAN. Code 6S is perhaps the most complete in the visible spectrum. Table 11.1 shows some characteristics of these physically based RTMs.

The use of RTMs is not limited to mere calculation of reflectance and radiances at ground level. Since they operate with a wide variety of parameters, they also allow the study of optics in many other fields of the science. The complexity of the actual internal physical solutions provided by these models is beyond the scope of this book.

Data input to RTMs Although a detailed description of the physical parameters involved in these models is outside of the scope of this book, a description of the type of input gives some indication on the effort required to run them.

The user must select:

- options that allow a more accurate calculation of molecular absorption in

the presence of multiple scattering;

- the type of atmospheric path;
- the kind of operating mode (output) desired;
- the kind of model atmosphere: a selection of standard atmospheres (explained below) or a user defined ones;
- the temperature and pressure profiles;
- the altitude profiles of water vapour, ozone, methane, nitrous oxide, carbon monoxide, and other gases;
- eventually the altitude and the sensor type.

RTM are relatively easy to use if the complexity of the atmospheric input is simplified by using one standard atmosphere as input.

Band transmission models adapted for RS image processing RTMs process extensive AC calculations and are not normally used for image processing. Instead, they are used on in individual cases (*eg*, to study a target of a certain characteristics), or for research studies where different RTM scenarios are constructed to assess the multitude of possible combinations of reflectance, viewing geometry and atmospheric conditions.

The results of those investigations can be used to build look-up tables (LUT databases), with which users can estimate the ground reflectance of a certain

target without actually involving a complex and time consuming RTM. This process is as accurate as the RTM used to build the LUT, and several times faster. It is used in models built especially for image processing, such as ATCOR ([24]), implemented in ERDAS Imagine. The advantage of these models is not only speed. The range of available options is very much focused on the problems routinely occurring during image pre-processing, for example where the atmospheric description is unknown for the image being processed.

An alternative approach conveniently built into the software is the possibility of working "backwards" from earth to satellite. Materials on the Earth have unique spectral signatures (see, *eg*, [15]) that are normally distorted by atmospheric interaction along the path towards the satellite. If the material of an observed feature is known (in a certain recognizable pixel), then the real spectral signature of this feature can be retrieved from a database of material reflectance, normally included in the software (see also Section 13.1). Due to atmospheric interferences the satellite reflectance does not match the database information. The model then provides tools to the user to modify the characteristics of the atmosphere within certain limits until the TOA reflectance matches the database reflectance. At this stage the user manages to "recreate" the atmosphere that produces the distortion. This atmosphere is applied to the entire image using the LUT in the software, performing a fast atmospheric correction without the need of measuring atmospheric components.

Other "simplified" RTMs for image processing still rely on the atmospheric description, but the load of necessary input information is reduced to a few standard parameters more widely measured. The algorithms inside these models

normally assume some restrictive assumptions (*ie*, due to the spatial resolution of the satellite) that allows a faster calculation, while keeping the error associated to these assumptions at tolerable values, normally reported in the software documentation.

As an example, SMAC ([23]) is a simplified version of Code 5S and 6S. It was originally designed for NOAA AVHRR data, and has been extended to include some high resolution satellites. It still requires information of ozone, aerosols and water vapour, but adapted to total amounts in vertical columns of atmosphere, while detailed profiles are not necessary. This information is more widely available, since it can be synthesized using sun photometers that are operated at ground stations in many countries around the world, and produce the necessary information on an hourly basis. See `http://www.cesbio.ups-tlse.fr/fr/serveurs4.htm` for access to SMAC.

Standard atmospheres Due to the rapid dynamics of the atmosphere in terms of temporal and spatial variation of its constituents, researchers have found the need to define some often-observed 'common profiles' corresponding to average atmospheric conditions in different parts of the Earth. Compilation of these 'fixed atmospheres' was based on actual radio soundings carried out at different research sites, resulting in so-called 'standard atmospheres', for example for mid-latitude summer, mid-latitude winter, tropical, desert, arctic, US standard, etc. Researchers use these well described standards to characterize the typical on-site atmospherics. RTMs have these standards built into the system, allowing the influence of different constituents to be compared under strict simulations. For instance, the influence of water vapour in the thermal part of the spectrum,

or of aerosols and air molecules in the visible part can be accurately predicted for different atmospheres, allowing sensitivity analysis to evaluate the importance of these constituents in attenuation processes at different wavelengths.

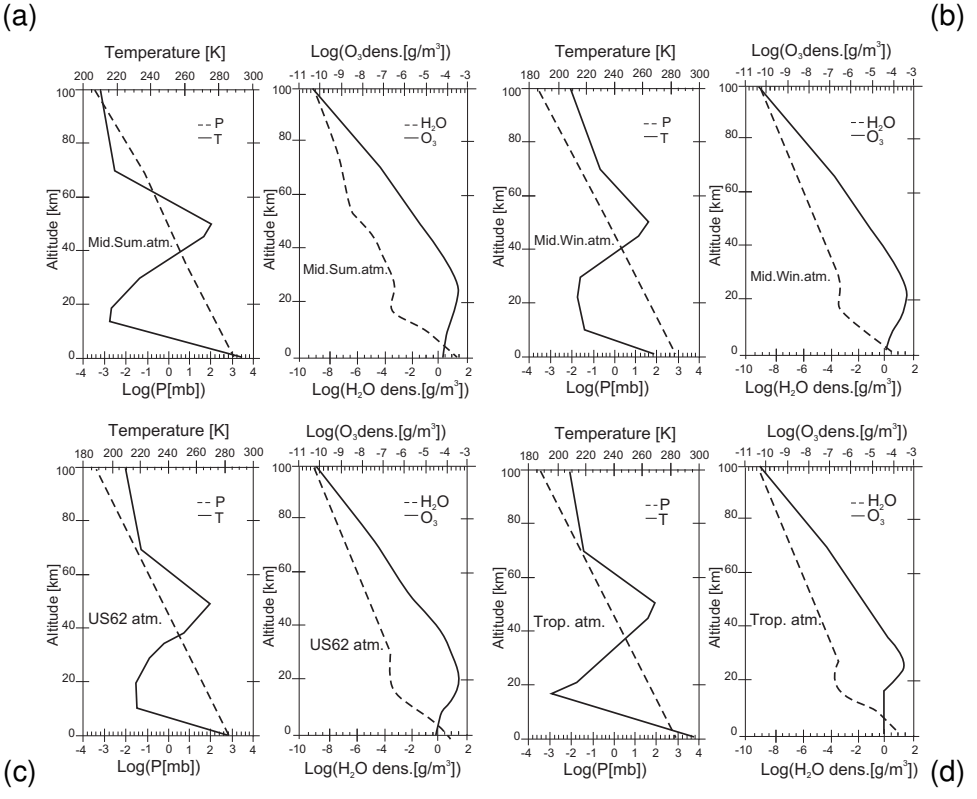


Figure 11.6: Model atmospheric profiles for mid-latitude summer (a), mid-latitude winter (b), US 62 standard (c) and tropical mode (d).

Figure 11.6 shows four standard atmosphere profiles. The left hand graph of

each figure shows the variation of temperature (in K) and pressure (in millibar) of the atmosphere up to an altitude of 100 km. The right hand graph shows the variation of ozone and water vapour with altitude on a logarithmic scale [g/cm^3]. Note that in general the pressure profile is similar in all atmospheres. Temperature profiles also have a similar shape, but the absolute values are very different, showing the importance of a good profile selection when analyzing atmospheric influences in thermal data. Ozone is mainly concentrated between 15 to 30 km, where most of the attenuation of the ultraviolet takes place. Water vapour is concentrated in the lower atmosphere in all cases, with a maximum close to the Earth's surface.

Summary

Radiometric corrections can be divided into relatively simple cosmetic rectifications, as well as atmospheric corrections. The cosmetic modifications are useful to reduce or compensate for data errors.

Atmospheric corrections constitute an important step in the pre-processing of RS data. Their effect is to re-scale the "raw radiance data" provided by the sensor to reflectance by correcting for atmospheric influence. They are important for generating image mosaics and for comparing multitemporal remote sensing data, but are also a critical prerequisite for the quantitative use of remote sensing data, for example to calculate surface temperatures (see Chapter 12), or to determine specific surface materials in spectrometer data (Chapter 13).

Following an overview of different techniques to correct data errors, absolute correction methods were explained. Focus was placed on radiative transfer models. Note that such corrections should be applied with care and only after understanding the physical principles behind these corrections.

Additional reading

[23]

Questions

The following questions can help to study Chapter 11.

1. Should radiometric corrections be performed before or after geometric corrections, and why?
2. 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?
3. Why is the selection of the appropriate standard atmosphere for a RTM critical?

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?

Chapter 12

Thermal remote sensing

12.1 Introduction

Thermal remote sensing is based on the measuring of electromagnetic radiation in the infrared region of the spectrum. Most commonly used are the intervals from 3 to 5 μm and 8 to 14 μm, in which the atmosphere is fairly transparent and the signal is only lightly attenuated by atmospheric absorption. Since the source of the radiation is the heat of the imaged surface itself (see Figures 2.8 and 2.18), the handling and processing of TIR data is considerably different from remote sensing based on reflected sunlight:

- The surface temperature is the main factor that determines the amount of energy that is radiated and measured in the thermal wavelengths. The temperature of an object varies greatly depending on time of the day, season, location, exposure to solar irradiation, etc and is difficult to predict. In reflectance remote sensing, on the other hand, the incoming radiation from the Sun is constant and can be readily calculated, although atmospheric correction has to be taken into account.
- In reflectance remote sensing the characteristic property we are interested in is the *reflectance* of the surface at different wavelengths. In thermal remote sensing, however, one property we are interested in is, how well energy is *emitted* from the surface at different wavelengths.
- Since thermal remote sensing does not depend on reflected sunlight, it can also be performed during the night (for some applications even better than during the day).

In section 12.2 the basic theory of thermal remote sensing is explained. Section 12.3 introduces the fundamental steps of processing TIR data to extract useful information. Section 12.4 illustrates in two examples how thermal remote sensing can be used in different application fields.

12.2 Principles of Thermal Remote Sensing

12.2.1 The physical laws

In Chapter 2 some radiation principles were introduced. From those we know that all objects above absolute zero temperature radiate EM energy. *Planck's radiation law* describes the amount of emitted energy per wavelength depending on the object's temperature:

Planck

$$M_{\lambda,T} = \frac{C_1}{\lambda^5\left[e^{(\frac{C_2}{\lambda T})} - 1\right]}, \tag{12.1}$$

where $M_{\lambda,T}$ is the spectral radiant emittance in (W m^{-3}), λ is the wavelength in (m), T is the absolute temperature in (K), C_1 is the first radiation constant, $3.74151 \cdot 10^{-16}$ (W m^2) and C_2 is the second radiation constant, 0.01438377 (mK).

Planck's radiation law is illustrated in Figure 12.1 for the approximate temperature of the Sun (about 6000 K) and the ambient temperature of the Earth's surface (about 300 K), respectively. These graphs are often referred to as black-body curves. The figure shows that for very hot surfaces (*eg*, the Sun), the peak of the black-body curve is at short wavelengths. For colder surfaces, such as the Earth, the peak of the black-body curve moves to lower wavelengths. This behaviour is described by *Wien's displacement law*:

Wien

$$\lambda_{max} = \frac{2898}{T}, \tag{12.2}$$

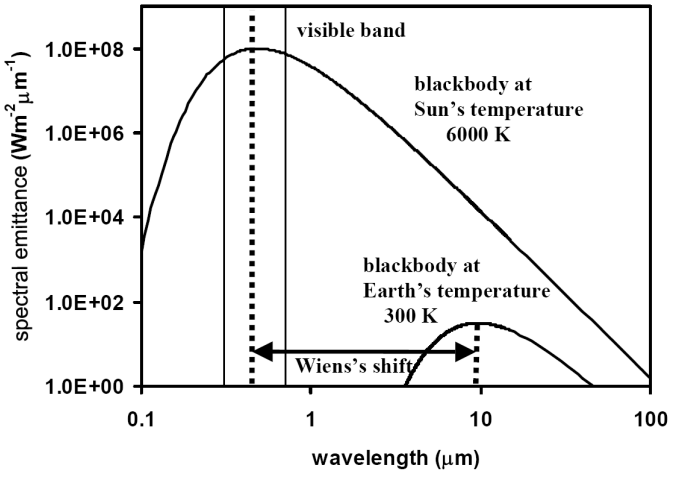


Figure 12.1: Illustration of Planck's radiation law for the Sun (6000 K) and the average Earth's surface temperature (300 K). Note the logarithmic scale for both x and y-axis. The dashed lines mark the wavelength with the emission maxima for the two temperatures.

where λ_{max} is the wavelength of the radiation maximum (in μm), T is the temperature (in K) and 2898 is a physical constant (in μm).

We can use Wien's law to predict the position of the peak of the black-body curve if we know the temperature of the emitting object. If you were interested in monitoring forest fires that burn at 1000 K, you could immediately turn to bands around 2.9 μm, where the radiation maximum for those fires is expected. For ordinary land surface temperatures around 300 K, wavelengths from 8 to 14 μm are most useful.

You can now understand why reflectance remote sensing (*ie*, based on reflected sunlight) uses short wavelengths in the visible and short-wave infrared, and

thermal remote sensing (based on emitted earth radiation) uses the longer wavelengths in the range 3 to 14 μm. Figure 12.1 also shows that the total energy (integrated area under the curve) is considerably higher for the Sun than for the cooler Earth's surface. This relationship between surface temperature and total radiant energy is known as the *Stefan-Boltzmann law*.

Stefan-Boltzmann

$$M = \sigma T^4, \tag{12.3}$$

where M is the total radiant emittance (W m^{-2}), σ is the *Stefan-Boltzmann constant*, $5.6697 \cdot 10^{-8}$ (W m^{-2} K^{-4}), and T is the temperature in (K).

The Stefan-Boltzmann law states that colder targets emit only small amounts of EM radiation. Wien's displacement law predicts that the peak of the radiation distribution will shift to longer wavelengths as the target gets colder. In Section 2.2.1 you have learnt that photons at long wavelengths have less energy than those at short wavelengths. Hence, in thermal RS we are dealing with a small amount of low energy photons, which makes their detection difficult. As a consequence of that, we often have to reduce spatial or spectral resolution when acquiring thermal data to guarantee a reasonable signal-to-noise ratio.

12.2.2 Black-bodies and emissivity

The three laws described above are, strictly speaking, only valid for an ideal radiator, which we refer to as a black-body. A black-body is a perfect absorber and a perfect radiator in all wavelengths. It can be thought of as a black object that reflects no incident EM radiation. But it is not only "black" in the visible bands, but in all wavelengths (of interest). If an object is a black-body, it behaves exactly as the theoretical laws predict. True black-bodies do not exist in nature, although some materials (*eg*, clean, deep water radiating between 8 to 12 μm) come very close.

Black-body

Materials that absorb and radiate only a certain fraction compared to a black-body are called grey-bodies. The fraction is a constant for all wavelengths. Hence, a grey-body curve is identical in shape to a black-body curve, but the absolute values are lower as it does not radiate as perfectly as a black-body.

Grey-body

A third group are the selective radiators. They also radiate only a certain fraction of a black-body, but this fraction changes with wavelength. A selective radiator may radiate perfectly in some wavelengths, while acting as a very poor radiator in other wavelengths. The radiant emittance curve of a selective radiator can then also look quite different from a ideal, black-body curve.

Selective radiator

The fraction of energy that is radiated by a material compared to a true black-body is also referred to as emissivity (ϵ_λ). Hence, emissivity is defined as:

Emissivity

$$\epsilon_\lambda = \frac{M_{\lambda,T}}{M^{BB}_{\lambda,T}}, \tag{12.4}$$

where $M_{\lambda,T}$ is the radiant emittance of a real material at a given temperature, and $M^{BB}_{\lambda,T}$ is a radiant emittance of a black-body at the same temperature.

Most materials are selective radiators. Their emissivity can change quite significantly with wavelength, and many materials have emissivity highs and lows at distinct wavelengths. Therefore, a spectral emissivity curve - also called 'emissivity spectrum' - in the thermal infrared can be used to determine the composition of an object similarly to how the spectral reflectance curve is used in the visible to short-wave infrared range. In fact, emissivity and reflectance for a given wavelength are also related to each other: objects that reflect well have a very poor ability to absorb/emit and vice versa. This behaviour is described in *Kirchhoff's law*, which is valid as long as the material is opaque.

Spectral emissivity

$$\epsilon_\lambda = 1 - \rho_\lambda, \tag{12.5}$$

where ρ_λ and ϵ_λ are reflectance (expressed as ratio) and emissivity, respectively, at a given wavelength λ.

As it is easier to measure reflectance than emissivity with laboratory instruments, Kirchhoff's law is often applied to calculate emissivity spectra from reflectance data rather than measure it directly. Note that the term 'broad-band

Emissivity spectra

emissivity' indicates the average emissivity value over a large part of the thermal spectrum, often from 8 to 14 μm. We should indicate this by writing ϵ_{8-14}. However, the symbol ϵ is often used without subscript.

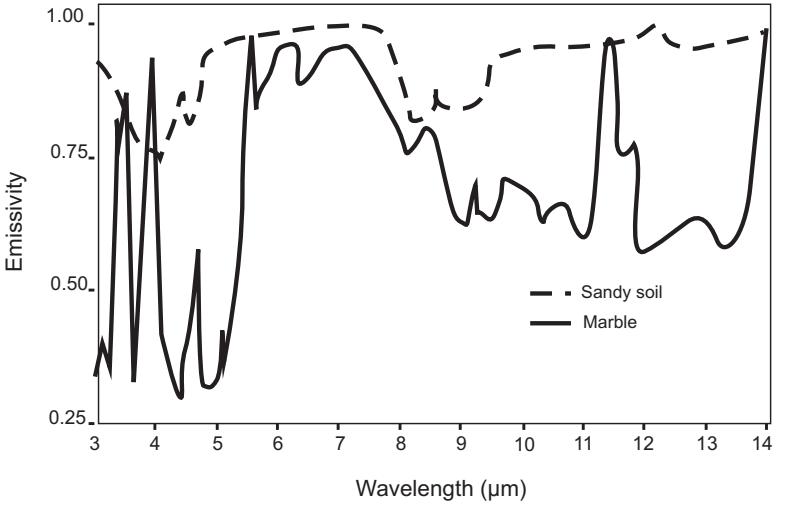


Figure 12.2: The TIR spectral emissivity curves of a sandy soil and a marble. Distinct emissivity highs and lows can be observed in the two spectral curves. From Johns Hopkins University Spectral Library.

12.2.3 Radiant and kinetic temperatures

The actual measurements done by a TIR sensor will relate to 'spectral radiance' (Wm^{-2} sr^{-1} μm^{-1}) that reaches the sensor for a certain wavelength band. We know that the amount of energy that is radiated from an object depends on its temperature and emissivity. That means that a cold object with high emissivity can radiate just as much energy as a considerably hotter object with low emissivity. We can decide to ignore the emissivity of the object. With the help of Planck's law we can directly calculate the ground temperature that is needed to create this amount of radiance in the specified wavelength band of the sensor if the object had a perfect emissivity of 1.0. The temperature calculated is the *radiant temperature* or T_{rad}. The terms 'brightness' or 'top-of-the-atmosphere' temperatures are also used frequently.

Radiant temperature

The radiant temperature calculated from the radiant energy emitted is in most cases smaller than the true, *kinetic temperature* (T_{kin}) that we could measure with a contact thermometer on the ground. The reason is that most objects have an emissivity lower than 1.0 and radiate incompletely. To calculate the true T_{kin} from the T_{rad}, we need to know or estimate the emissivity. The relationship between T_{kin} and T_{rad} is:

Kinetic temperature

$$T_{rad} = \epsilon^{1/4} T_{kin}. \tag{12.6}$$

With a single thermal band (*eg*, Landsat-7 ETM+), ϵ has to be estimated from other sources. One way is to do a land cover classification with all available bands and then assign an ϵ value for each class from an emissivity table (*eg*, 0.99 for water, 0.85 for granite).

Separation of emissivity and temperature

In multispectral TIR, several bands of thermal wavelengths are available. With emissivity in each band as well as the surface temperature (T_{kin}) unknown, we still have an underdetermined system of equations. For that reason, it is necessary to make certain assumptions about the shape of the emissivity spectrum we are trying to retrieve. Different algorithms exist to separate the influence of temperature from the emissivity. We will look at this issue in more detail in the following section.

12.3 Processing of thermal data

As mentioned in the introduction, many processing steps of thermal data are different from those in the visual, NIR and SWIR regions. The different processing steps strongly depend on the application the thermal data. In the following sections we will deal with some examples. These can be roughly classified in two categories: those for which the emissivity values are most important, because these can be used to distinguish between geological surface characteristics, and those for which the actual surface temperature needs to be determined. For some applications simple image enhancement techniques are sufficient.

12.3.1 Band ratios and transformations

For some applications, image enhancement of the thermal data is sufficient to achieve the necessary outputs. They can be used, for example, to delineate relative differences in surface emissivity or surface temperature.

Band ratios

By ratioing two bands we can highlight the areas where a surface material of interest is predominant. In order to do so, one would ratio two bands near a rapid change in the emissivity spectrum of that material (*eg*, low emissivity around 9 μm in sandstone). We can consult spectral libraries to find out where we can expect these sharp changes in the emissivity spectrum of a particular material (see Figure 12.2). Band ratios also reduce the influence of differences in surface temperature. This can be an advantage if the study area is affected by differential heating (hill slopes).

Transformations

In a multispectral, thermal data set each band shows the influence of the emissivities of the surface materials as well as of the surface temperature. As the surface temperature does not vary with wavelength, a large percentage of the information contained in the images is identical in all bands; they are said to be correlated. By applying an image transformation, such as Principal Component Analysis (PCA) or 'decorrelation stretching', we can minimize the common information (*ie*, surface temperature) and enhance the visibility of the differences in the bands.

12.3.2 Determining kinetic surface temperatures

Instead of recording spectral radiances, the TIR sensor records DNs. In the case of Landsat the digital numbers range from 0 to 255 (8 bits), whereas in the case of NOAA/AVHRR the digital numbers use 10 bits per pixel and ASTER uses 12 bits per pixel in the TIR bands. Therefore, for most satellites the DN-values need to be converted to radiances first. This is usually done through the use of gain and offset values for the linear relationship, which are always listed in the appropriate satellite manuals and websites.

DN to radiance

The next step is to convert a radiance into radiant temperature at the top-of-the-atmosphere (TOA) using equations such as 12.7:

$$T = \frac{K_2}{ln\left[\frac{K_1}{L_\lambda} + 1\right]}, \tag{12.7}$$

where, for example, the Landsat-7 ETM+ constants are K_1=666.09 Wm^{-2} sr^{-1} μm^{-1} and K_2=1282.71 K.

Radiance to temperature

The satellite "sees" only the radiation that emerges from the atmosphere, L^{sat}, that is the radiance leaving the surface multiplied by the transmittance, plus the radiance produced by the atmosphere. The radiance leaving the surface, $L\uparrow$, is a combination of the downwelling atmospheric radiance, $L\downarrow$, and the radiance produced by the surface of the Earth. Remember that the Earth's surface is not a perfect black-body, but usually a selective radiator with an emissivity less than

Radiance at the sensor

1. Therefore, part of the downwelling radiance is reflected. See also Kirchoff's Law (12.5). The radiance from the Earth is given as the emissivity, ϵ, multiplied by the black-body radiance, L^{BB} . These considerations are usually combined in the following equation:

$$L^{sat} = \left[\epsilon \cdot L^{BB} + (1-\epsilon)L\downarrow\right]\tau + L\uparrow \tag{12.8}$$

In summary, L^{sat} is measured by the satellite, but we need L^{BB} to determine the kinetic surface temperature. Equation 12.8 shows that in practice there are two problems in the processing of thermal images:

- determination of the upwelling ($L\uparrow$) and downwelling ($L\downarrow$) atmospheric radiances, together with the atmospheric transmittance (τ) in the thermal range of the EM spectrum;
- determination of the emissivities and surface temperature.

The first of these problems requires knowledge of atmospheric parameters during the time of satellite overpass. Once these are known, software such as MODTRAN4 (see Section 2.5) can be applied to produce the required parameters $L\uparrow$, $L\downarrow$ and τ.

Because the emissivities are wavelength dependent, equation 12.8 leads to n equations with $n+1$ unknowns, where n is the number of bands in the thermal image. Additional information is, therefore, required to solve the set of equations. Most methods make use of laboratory derived information with regard to

the shape of the emissivity spectra. This process is called temperature-emissivity separation. The algorithms, however, are rather complex and outside the scope of this chapter. A complete manual on thermal processing, including examples and more details on the mathematics involved, is available at ITC's Department of Water Resources.

12.4 Thermal applications

This section provides two example applications, for which thermal data can be used. In general the applications of thermal remote sensing can be divided into two groups:

- The main interest is the study of the surface composition by looking at the surface emissivity in one or several wavelengths.
- The focus is on the surface temperature, its spatial distribution or change over time.

12.4.1 Rock emissivity mapping

As we have seen in Figure 12.2, many rock and soil types show distinct spectra in the thermal infrared. The absorption bands are mainly caused by silicate minerals, such as quartz or feldspars, that make up a large percentage of the World's rocks and soils. By carefully studying thermal emissivity spectra, we can identify different mineral components the target area is composed of. Figure 12.3 shows a thermal image taken by the MASTER airborne sensor over an area near Cuprite, Nevada. The original colour composite was decorrelation stretched for better contrast. It clearly shows the different rock units in the area. The labels show a limestone (lst) in tints of light green in the south, volcanic tuff (t) in cyan colour in the north. A silica (s) capped hill shows shades of dark orange near the centre of the image. Several additional units can also be distinguished, based on this band combination alone.

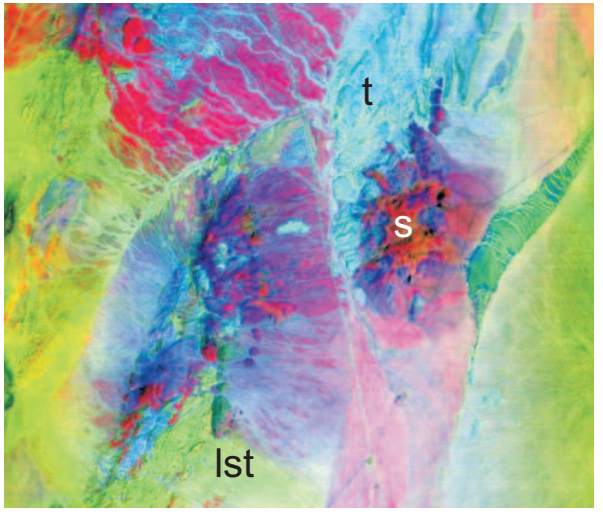


Figure 12.3:
Decorrelation stretched colour composite of a MASTER image; RGB = Bands 46, 44, 42; see text for more information on rock units. Scene is 10 km wide.

12.4.2 Thermal hotspot detection

Another application of thermal remote sensing is the detection and monitoring of small areas with thermal anomalies. The anomalies can be related to fires, such as forest fires or underground coal fires, or to volcanic activity, such as lava flows and geothermal fields. Figure 12.4 shows an ASTER scene that was acquired at night. The advantage of night images is that the sun does not heat up the rocks surrounding the anomaly, as would be the case during the day. This results in better contrast between the anomaly temperatures themselves and the surrounding rocks. This particular image was taken over the Wuda coal-mining area in China in September 2002. Hotter temperatures are reflected by brighter shades of grey. On the right side the yellow river is clearly visible, since water does not cool down as quickly as the land surface does, due to thermal inertia. Inside the mining area (fine, white box), several hotspots are visible with elevated temperatures compared to the surrounding rocks. The inset shows the same mining area slightly enlarged. The hottest pixels are coloured in orange and show the locations of coal fires. If images are taken several weeks or even years apart, the development of these underground coal fires as well as the effect of fire fighting efforts can be monitored quite effectively with thermal remote sensing.

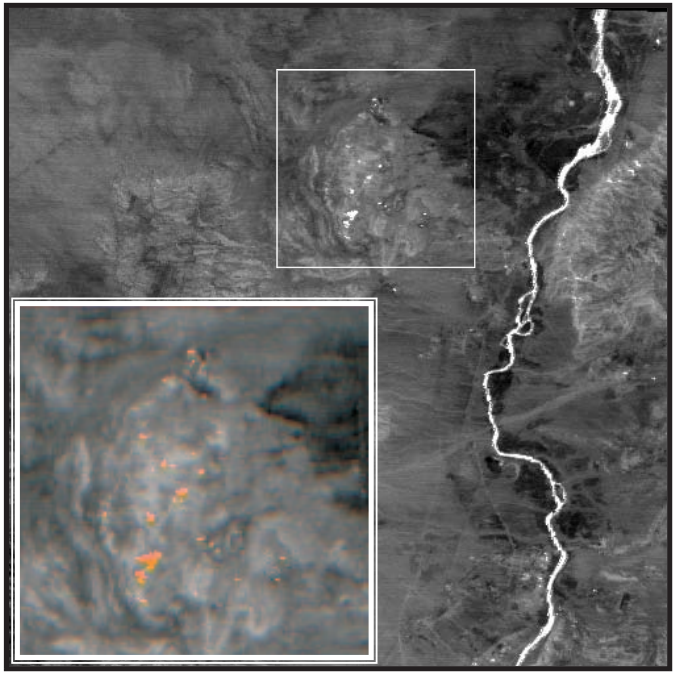


Figure 12.4: ASTER thermal band 10 over Wuda, China. Light coloured pixels inside the mining area (fine, white box) are caused mainly by coal fires. Inset: pixels exceeding the background temperature of 18° C are coloured in orange for better visibility of the fire locations. Scene is approximately 45 km wide.

Summary

This chapter has provided an introduction to thermal remote sensing, a passive technique that is aimed at recording radiation emitted by the material or surface of interest. It was explained how thermal RS is mostly applied to the mid-IR and thermal IR, and how the amount and peak wavelengths of the energy emitted is a function of the object's temperature. This explained why reflectance remote sensing (*ie*, based on reflected sunlight) uses short wavelengths (in the visible to SWIR range), whereas thermal remote sensing uses the longer wavelengths. In addition to the basic physical laws, the concepts of black-body radiation and emissivity were explained. Incomplete radiation, *ie*, a reduced emissivity was shown to account for the kinetic temperatures often being lower than the corresponding radiant temperature.

The subsequent section on the processing of thermal data gave an overview of techniques aimed at the differentiation of surface materials, as well as methods to calculate actual surface temperatures. The last section provided examples of how the surface distribution of different rock types can be mapped, and how thermal anomalies can be assessed. The methods are applicable to many different problems, including coal-fire mapping, sea surface temperature monitoring, weather forecasting, but also in the search-and-rescue of missing persons.

Questions

The following questions can help to study Chapter 12.

1. What is the total radiant energy from an object at a temperature of 300 K?
2. Calculate the wavelength of peak energy from a volcanic lava flow of about 1200° C.
3. Is the kinetic temperature higher than the radiant temperature or the other way around? Explain your answer with an example.
4. For a Landsat-7 ETM+ image a certain pixel has a spectral radiance of 10.3 Wm^{-2} sr^{-1} μm^{-1}. Determine its radiant temperature.
5. For a sea level summer image (Landsat-5), the following atmospheric parameters are determined with MODTRAN4:

   - The atmospheric transmittance τ is 0.7.
   - The upwelling radiance $L\uparrow$ is 2.4 Wm^{-2} sr^{-1} μm^{-1}.
   - The downwelling radiance $L\downarrow$ is 3.7 Wm^{-2} sr^{-1} μm^{-1}.

   The broad-band surface emissivity ϵ is 0.98 and the radiance L_{sat} observed at the satellite is 8.82 Wm^{-2} sr^{-1} μm^{-1}.

   - Calculate the black-body radiance L^{BB}.
   - Determine the surface radiant temperature and the kinetic temperature.


6. How can you enhance thermal images for visual interpretation?

7. Given a satellite sensor with multiple thermal bands, can you determine the emissivities? What are the difficulties?

Chapter 13

Imaging Spectrometry

13.1 Introduction

You have learnt in Chapter 2 that materials of interest may be discriminated based on their spectral reflectance curves (*eg*, Figure 2.16). In this chapter we will call spectral reflectance curves 'reflectance spectra'. Most multispectral sensors that were discussed in Chapter 4 acquire data in a number of relatively broad wavelength bands. However, typical diagnostic absorption features, characterizing materials of interest in reflectance spectra, are in the order of 20 nm to 40 nm in width. Hence, the broadband sensors under-sample this information and do not allow to exploit the full spectral resolving potential available. Imaging spectrometers typically acquire images in a large number of spectral bands (more than 100). These bands are narrow (less than 10 nm to 20 nm in width), and contiguous (*ie*, adjacent), which enables the extraction of reflectance spectra at pixel scale (see Figure 13.1). Such narrow spectra enable the detection of the diagnostic absorption features. Different names have been coined for this field of remote sensing, including imaging spectrometry, imaging spectroscopy and hyperspectral imaging.

Figure 13.2 illustrates the effect of spectral resolution on the mineral kaolinite. From top to bottom the spectral resolution increases from 100–200 nm (Landsat), 20–30 nm (GERIS), 20 nm (HIRIS), 10 nm (AVIRIS), to 1–2 nm (USGS laboratory spectrum). With each improvement in spectral resolution, the diagnostic absorption features and, therefore, the unique shape of kaolinite's spectrum, become more apparent.

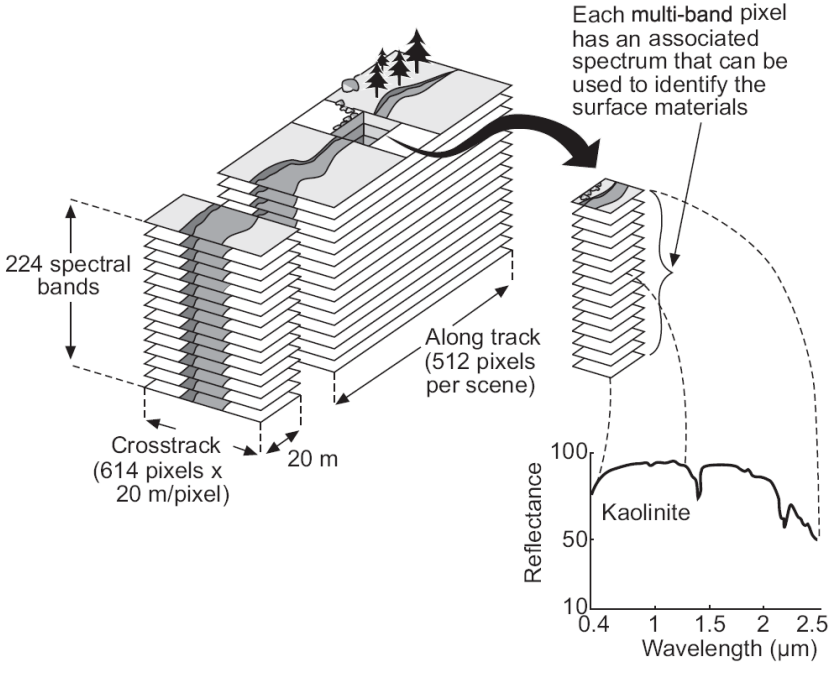


Figure 13.1: Concept of imaging spectrometry (adapted from [32]).

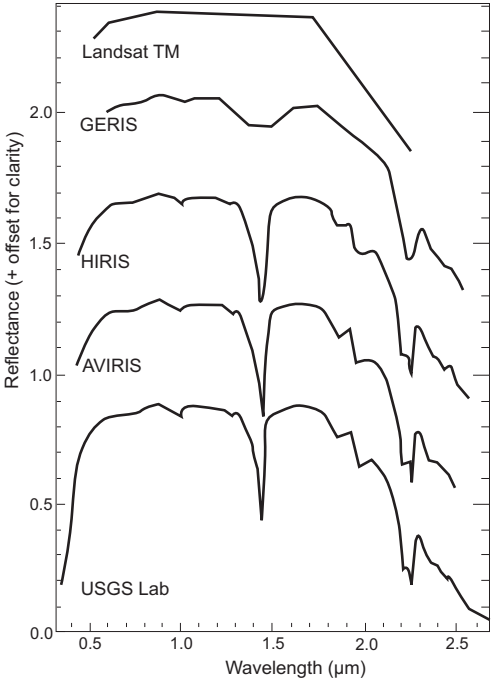


Figure 13.2: Example of a kaolinite spectrum at the original resolution (source: USGS laboratory) and at the spectral resolutions of various imaging devices. Note that spectrum are progressively offset upward by 0.4 units for clarity (adapted from USGS).

13.2 Reflection characteristics of rocks and minerals

Rocks and minerals reflect and absorb electromagnetic radiation as a function of the wavelength of the radiation. Reflectance spectra show these variations in reflection and absorption for various wavelengths (Figure 13.3). By studying the reflectance spectra of rocks, individual minerals and groups of minerals may be identified. In earth sciences, absorption in the wavelength region from 0.4 μm to 2.5 μm is commonly used to determine the mineralogical content of rocks. In this region various groups of minerals have characteristic reflectance spectra, for example phyllo-silicates, carbonates, sulphates, and iron oxides and hydroxides. High-resolution reflectance spectra for studying mineralogy can easily be obtained in the field or in a laboratory using field spectrometers.

Reflectance spectra

Processes that cause absorption of electromagnetic radiation occur at a molecular and atomic level. Two types of processes are important in the 0.4 μm to 2.5 μm range: electronic processes and vibrational processes ([5]). In electronic processes, individual atoms or ions in minerals absorb photons of specific wavelengths, which cause absorptions at certain wavelengths in reflectance spectra. An example is absorption by Fe^{3+} atoms in iron oxides and hydroxides, which gives these minerals a red colour. In vibrational processes, molecular bonds absorb photons, which results in vibration of these molecules. Examples of bonds that absorb radiation are Al-OH bonds in clay minerals, bonds in H_2O and OH^- in hydrous minerals, and in $CO_3{}^{2-}$ in carbonate minerals.

Electronic processes

Vibrational processes

Reflectance spectra respond closely to the crystal structure of minerals and can be used to obtain information of crystallinity and chemical composition of min-

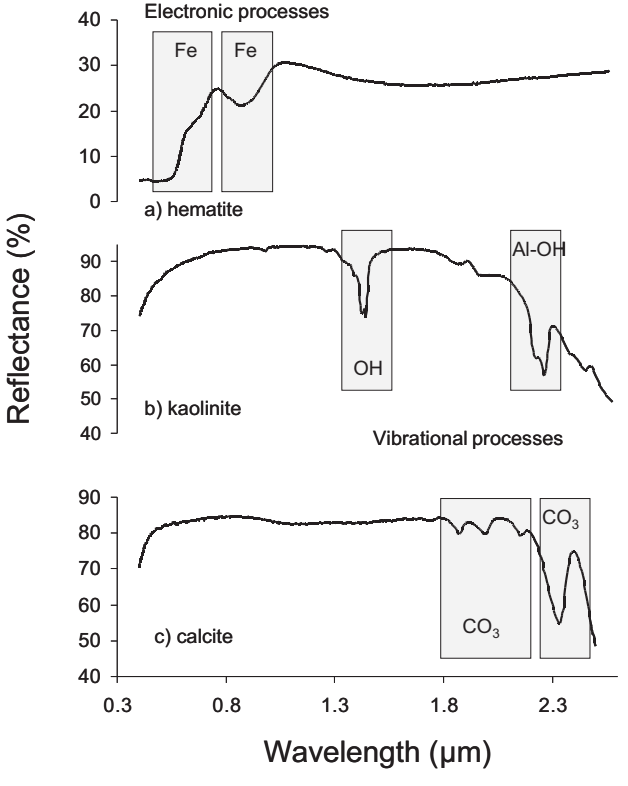


Figure 13.3: Effects of electronic and vibrational processes on absorption of electromagnetic radiation.

erals.

13.3 Pre-processing of imaging spectrometer data

Pre-processing of imaging spectrometer data involves radiometric calibration (see Chapter 11), which provides transfer functions to convert DN-values to at-sensor radiance. The at-sensor radiance data have to be corrected by the user for atmospheric effects to obtain surface reflectance data. In Chapter 11 an overview is given of the use of radiative transfer models for atmospheric correction. The correction provides absolute reflectance data, because the atmospheric influence is modelled and removed.

Radiometric calibration
Atmospheric correction

Alternatively, users can perform a scene-dependent relative atmospheric correction using empirically derived models for the radiance-reflectance conversion, using calibration targets found in the imaging spectrometer data set. Empirical models that are often used include what are called the flat-field correction and the empirical-line correction. Flat-field correction achieves radiance-reflectance conversion by dividing the whole data set on a pixel-by-pixel basis by the mean value of a target area within the scene that is spectrally and morphologically flat, spectrally homogeneous and has a high albedo. Conversion of raw imaging spectrometer data to reflectance data using the empirical-line method requires the selection and spectral characterization (in the field with a spectrometer) of two calibration targets (a dark and a bright target). This empirical correction uses a constant gain and offset for each band to force a best fit between sets of field spectra and image spectra that characterize the same ground areas, thus removing atmospheric effects, residual instrument artefacts and viewing geometry effects.

Relative correction

13.4 Atmospheric correction of imaging spectrometer data

Spectral radiance curves of uncorrected imaging spectrometer data have the general appearance of the solar irradiance curve, with radiance values decreasing towards longer wavelengths, and exhibiting several absorption bands due to scattering and absorption by gasses in the atmosphere.

The effect of atmospheric calibration algorithms is to re-scale the "raw radiance data" provided by imaging spectrometers to reflectance by correcting for atmospheric influence. The result is a data set in which each multi-band pixel is represented by a reflectance spectrum that can be directly compared to reflectance spectra of rocks and minerals acquired either in the field or in the laboratory (see Figure 13.1). Reflectance data obtained can be absolute radiant energy or apparent reflectance relative to a certain standard in the scene. Calibration to reflectance can be conducted to result in absolute or relative reflectance data. See Chapter 11 for more information on atmospheric correction.

13.5 Thematic analysis of imaging spectrometer data

Once reflectance-like imaging spectrometer data are obtained, the logical next step is to use diagnostic absorption features to determine and map variations in surface composition. New analytical processing techniques have been developed to analyze such high-dimensional spectral data sets. These methods are the focus of this section. Such techniques can be grouped into two categories: *spectral matching* approaches and *subpixel classification* methods.

13.5.1 Spectral matching algorithms

Spectral matching algorithms aim at quantifying the statistical or physical relationship between measurements at a pixel scale and field or laboratory spectral responses of target materials of interest. A simple spectral matching algorithm used is 'binary encoding', in which an imaged reflectance spectrum is encoded as

Binary encoding

$$h_i = \begin{cases} 0 & \text{if } x_i \leq T \\ 1 & \text{if } x_i > T \end{cases} \tag{13.1}$$

where x_i is the brightness value of a pixel in the ith channel, T is the user specified threshold (often the average brightness value of the spectrum is used for T), and h_i is the resulting binary code for the pixel in the ith band.

This binary encoding provides a simple mean of analyzing data sets for the presence of absorption features, which can be directly related to similar encoding profiles of known materials.

An often used spectral matching technique in the analysis of imaging spectrometer data sets is the so-called spectral angle mapper (SAM). In this approach, the spectra are treated as vectors in a space with a dimensionality equal to the number of bands, n. SAM calculates the spectral similarity between an unknown reflectance spectrum, $\vec{t}$ (consisting of band values t_i), and a reference (field or

SAM

laboratory) reflectance spectrum, $\vec{r}$ (consisting of band values r_i), and expresses the spectral similarity of the two in terms of the vector angle, θ, between the two spectra as calculated using all the bands. In vector notation,

$$\theta = \cos^{-1}\left[\frac{\vec{t} \bullet \vec{r}}{\|\vec{t}\| \cdot \|\vec{r}\|}\right] \tag{13.2}$$

or, using the band notation, as

$$\theta = \cos^{-1}\left[\frac{\sum_{i=1}^{n} t_i r_i}{\sqrt{\sum_{i=1}^{n} t_i^2 \sum_{i=1}^{n} r_i^2}}\right]. \tag{13.3}$$

The outcome of the spectral angle mapping for each multi-band pixel is an angular difference, measured in radians ranging from zero to $\pi/2$, which gives a qualitative measure for comparing known and unknown spectra. The smaller the angle, the more similar the two spectra.

13.5.2 Spectral unmixing

Often the electromagnetic radiation observed as reflectance values results from the spectral mixture of a number of ground classes present at the sensed surface. Researchers have shown that spectral mixing can be considered a linear process if:

1. multiple scattering does not occur inside the same material, *ie*, no multiple bounces occur,
2. no interaction between materials occurs, *ie*, each photon sees only one material, and
3. the scale of mixing is very large as opposed to the grain size of the materials.

Various sources contribute to spectral mixing:

1. optical imaging systems integrate reflected light from each ground resolution cell,
2. all materials present in the IFOV contribute to the mixed reflectance sensed for a ground resolution cell, and
3. variable illumination conditions due to terrain relief effects result in spectrally mixed signals.

In general, there are 5 types of spectral mixtures:

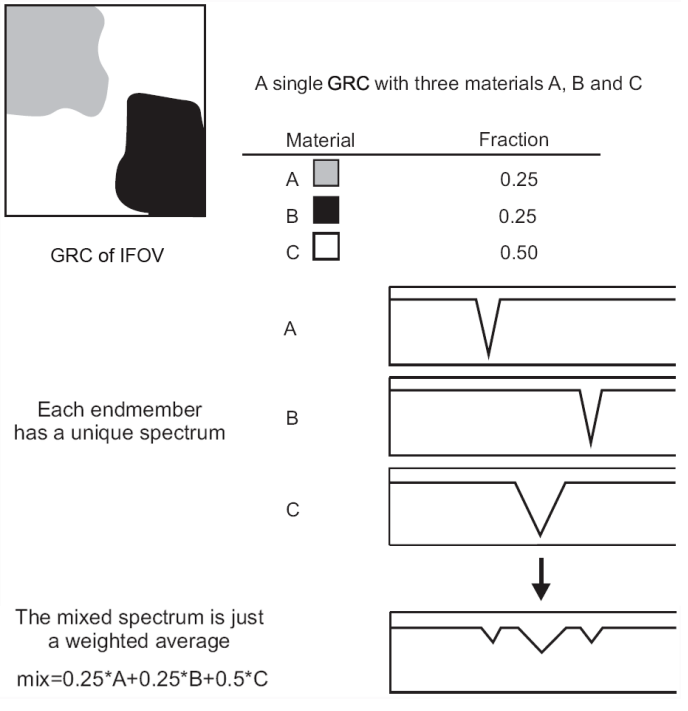


Figure 13.4: Concept of signal mixing and spectral unmixing.

1. Linear Mixture. The materials in the field of view are optically separated so there is no multiple scattering between components. The combined signal is simply the sum of the fractional area times the spectrum of each component. This is also called area mixture (see Figure 13.4).

2. Intimate Mixture. An intimate mixture occurs when different materials are in close contact in a scattering surface, such as the mineral grains in a soil or rock. Depending on the optical properties of each component, the resulting

signal is a highly non-linear combination of the end-member spectra.

3. Coatings. Coatings occur when one material coats another. Each coating is a scattering/transmitting layer whose optical thickness varies with material properties and wavelength. Most coatings yield truly non-linear reflectance properties of materials.

4. Molecular Mixtures. Molecular mixtures occur on a molecular level, such as two liquids, or a liquid and a solid mixed. Reflection is non-linear.

5. Multiple Scattering. Multiple reflections occur within the same material or between different materials. With multiple reflections, the reflection spectra are multiplied rather than added, as occurs due to the integration at the sensor. The multiplication results in a non-linear mixture.

The type and nature of the mixing systematics are crucial to the understanding of the mixed signal. In many processing approaches, linearity is assumed. This is also the case in spectral unmixing. From the list above, it can be observed that the assumption of linearity is only true in few cases.

Rather than aiming at representing the terrain in terms of a number of fixed classes, as is done in digital image classification (see Chapter 8), spectral unmixing (see Figure 13.4) acknowledges the compositional nature of natural surfaces. Spectral unmixing strives at finding the relative or absolute fractions (or abundance) of a set of spectral components that together contribute to the observed reflectance of ground resolution cells. The spectral components of the set are known reference spectra, which, in spectral unmixing, we call *end-members*. The outcome of such analysis is a new set of images that for each selected end-

Fractions of end-members

member portray the fraction of this end-member spectrum within the total spectrum of the cell. Mixture modelling is the forward process of deriving mixed signals from pure end-member spectra, whereas spectral unmixing aims at doing the reverse, deriving the fractions of the pure end-members from the mixed pixel.

A linear combination of spectral end-members is chosen to decompose the mixed reflectance spectrum of each cell, $\vec{R}$, into fractions of its end-members, $\vec{R}e_j$, by

$$\vec{R} = \sum_{j=1}^{n} f_j \vec{R}e_j + \vec{\epsilon} \text{ and } 0 \leq \sum_{j=1}^{n} f_j \leq 1 \tag{13.4}$$

where $\vec{R}$ is the reflectance of the mixed spectrum of each cell, f_j is the fraction of each end-member j, $\vec{R}e_j$ is the reflectance of the end-member spectrum j, j indicates each of the n end-members, and $\vec{\epsilon}$ is the residual error, *ie*, the difference between the measured and modelled DN-values. A unique solution is found from this equation by minimizing the residual error, $\vec{\epsilon}$, in a least-squares method.

Subpixel classification

13.6 Applications of imaging spectrometry data

In this last section, a brief outline of current applications in various fields relevant to the thematic context of ITC are given. These include examples in the area of geologic mapping and resource exploration, vegetation sciences, and hydrology.

13.6.1 Geology and resources exploration

Imaging spectrometry is used in an operational mode by the mineral industry for surface mineralogy mapping to aid in ore exploration. Other applications of the technology include lithological and structural mapping. The petroleum industry is developing methods for the implementation of imaging spectrometry at a reconnaissance stage as well. The main targets are hydrocarbon seeps and microseeps.

Other application fields include environmental geology (and related geobotany), in which currently much work is done on acid mine drainage and mine waste monitoring. Atmospheric effects resulting from geologic processes, as for example, the prediction and quantification of various gases in the atmosphere, such as sulfates emitted from volcanoes for hazard assessment, is also an important field. In soil science, much emphasis has been placed on the use of spectrometry for soil surface properties and soil compositional analysis. Major elements, such as iron and calcium, as well as cation-ion exchange capacity, can be estimated from imaging spectrometry. In a more regional context, imaging spectrometry has been used to monitor agricultural areas (per-lot monitoring) and semi-natural areas. Recently, spectral identification from imaging spectrometers has been successfully applied to mapping of swelling clays minerals smectite, illite, and kaolinite in order to quantify the swelling potential of expansive soils. It should be noted that mining companies, and to a lesser extent petroleum companies, are operationally exploiting imaging spectrometer data for reconnaissance-level exploration campaigns.

13.6.2 Vegetation sciences

Much research in vegetation studies has emphasized on leaf biochemistry and structure, and canopy structure. Biophysical models for leaf constituents are currently available, as are soil-vegetation models. Estimates of plant material and structure and biophysical parameters include: carbon balance, yield/volume, nitrogen, cellulose, chlorophyll, etc. The leaf area index and vegetation indices have been extended to the hyperspectral domain and remain important physical parameters for characterizing vegetation. One ultimate goal is the estimation of biomass and the monitoring of changes therein. Several research groups investigate the bidirectional reflectance function (see Appendix A) in relation to vegetation species analysis and floristics. Vegetation stress by water deficiency, pollution sources (such as acid mine drainage), and geobotanical anomalies in relation to ore deposits or petroleum and gas seepage, links vegetation analysis to exploration. Another upcoming field is precision agriculture, in which imaging spectrometry aids in better agricultural practices. An important factor in vegetation health status is the chlorophyll absorption and, in relation to that, the position of the red edge determined using the red-edge index. The red edge is the name given to the steep increase in the reflectance spectrum of vegetation, between the visible red and the near-infrared wavelengths.

13.6.3 Hydrology

In hydrological sciences, the interaction of electromagnetic radiation with water, and the inherent and apparent optical properties of water are a central issue. Very important in imaging spectrometry of water bodies is the atmospheric correction and air-water interface corrections. Water quality of freshwater aquatic environments, estuarine environments and coastal zones are of importance to national water bodies. Detection and identification of phytoplankton-biomass, suspended sediments and other matter, coloured dissolved organic matter, and aquatic vegetation (*ie*, macrophytes) are crucial parameters in optical models of water quality. Much emphasis has been put on the mapping and monitoring of the state and the growth or brake-down of coral reefs, as these are important in the CO_2 cycle. In general, many multi-sensor missions such as Terra and Envisat are directed towards integrated approaches for global change studies and global oceanography. Atmosphere models are important in global change studies and aid in the correction of optical data for scattering and absorption due to atmospheric trace gasses. In particular, the optical properties and absorption characteristics of ozone, oxygen, water vapor, and other trace gasses, and scattering by molecules and aerosols are important parameters in atmosphere studies. All these can be and are derived from imaging spectrometers.

13.7 Imaging spectrometer systems

In Chapter 4 an overview of spaceborne multispectral and hyperspectral sensors was given. Here we provide a short historic overview of imaging spectrometer sensor systems with examples of presently operational airborne and spaceborne systems. The first civil scanning imaging spectrometer was the Scanning Imaging Spectroradiometer (SIS) constructed in the early 1970's for NASA's Johnson Space centre. After that, civil airborne spectrometer data were collected in 1981 using a one-dimensional profile spectrometer developed by the Geophysical Environmental Research Company, which acquired data in 576 channels covering the 0.4 to 2.5 μm wavelength range, followed by the Shuttle Multispectral Infrared Radiometer (SMIRR) in 1981. The first imaging device was the Fluorescence Line Imager (FLI, also known as the Programmable Multispectral Imager, PMI), developed by Canada's Department of Fisheries and Oceans in 1981. This was followed by the Airborne Imaging Spectrometer (AIS), developed at the NASA Jet Propulsion Laboratory, which was operational from 1983 onward acquiring 128 spectral bands in the range of 1.2 to 2.4 μm. The FOV of 3.7° resulted in 32 pixels across-track. A later version of the instrument, AIS-2, covered the 0.8–2.4 μm region, acquiring images 64 pixels wide. Since 1987, NASA is operating the successor of the AIS systems, AVIRIS, the Airborne Visible/Infrared Imaging Spectrometer. The AVIRIS scanner collects 224 contiguous bands resulting in a complete reflectance spectrum for each 20 m by 20 m GRC in the 0.4 μm to 2.5 μm region, with a sampling interval of $<$10 nm. The FOV of the AVIRIS scanner is 30 degrees, resulting in a swath of 10.5 km from 20 km altitude. AVIRIS uses scanning optics and four spectrometers to image a 614 pixel swath simultaneously in 224 contiguous spectral bands over the 400 nm to 2500 nm wavelength range. Table 13.1 provides examples of some currently

operational airborne imaging spectrometer systems.

The first civil earth observing imaging spectrometer was the LEWIS satellite, carrying the Hyperspectral Imager, from the TRW company. Lewis was launched in 1997, but failed after 3 days in orbit. It was intended to cover the 0.4–1.0 μm range with 128 bands and the 0.9–2.5 μm range with 256 bands of 5 nm and 6.25 nm bandwidth, respectively.

On-board Terra, there are two spaceborne imaging spectrometer systems: the US-Japanese ASTER and the MODIS. Terra was introduced in Chapter 4. In 2000, NASA launched the EO-1 satellite carrying the Hyperion imaging spectrometer. EO-1 was introduced in Chapter 4.

The European Space Agency is currently operating MERIS on the ENVISAT platform. MERIS is a fully programmable imaging spectrometer. Envisat-1 was introduced in Chapter 4.

Instrument	Manufacturer	Spectral range (nm)	Bands	Bandwidth (nm)
Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)	NASA (US)	400–2500	224 bands	10
The Airborne Imaging Spectrometer (AISA)	SPECIM (Finland)	430–900	288 bands	1.63–9.8
Compact Airborne Spectrographic Imager (CASI)	ITRES (Canada)	400–870	288 bands	1.9
Digital Airborne Imaging Spectrometer (DAIS)	Geophysical Environmental Research Corporation (US)	500–12300	79 bands	15–2000
Hyperspectral Mapper (HyMAP)	Integrated Spectronics (Australia)	400–2500	126 bands	10–20
Multispectral Infrared and Visible Imaging Spectrometer (MIVIS)	SenSyTech, Inc. (US)	430–12700	102 bands	8–500
Reflective Optics System Imaging Spectrometer (ROSIS)	Dornier Satellite Systems (Germany)	430–850	84 bands	5

Table 13.1: Examples of some operational airborne imaging spectrometer systems.

Summary

Chapter 13 has given an overview of the concepts and methods of imaging spectrometry. It was explained how, on the one hand, it is similar to multispectral remote sensing in that a number of visible and NIR bands are used to study the characteristics of a given surface. However, imaging spectrometers typically acquire data in a much larger number of narrow and contiguous spectral bands. This makes it possible to extract "per-pixel reflectance spectra", which are useful for the detection of the diagnostic absorption features that allow us to determine and map variations in surface composition.

The section on pre-processing showed that for some applications a scene dependent relative atmospheric correction is sufficient. However, it was also explained when an absolute radiometric calibration must be applied, which provides transfer functions to convert DN-values to at-sensor radiance.

Once the data have been corrected, spectral matching and subpixel classification methods can be used to relate observed spectra to known spectral responses of different materials, or to identify which materials are present within a single ground resolution cell. Examples were provided in the areas of geologic mapping and resource exploration, vegetation sciences, and hydrology.

Additional reading

[5], [10], and [32].

Questions

The following questions can help to study Chapter 13.

1. What are the advantages and disadvantages of hyperspectral remote sensing in comparison with multispectral Landsat-type scanning systems?
2. In which part of the electromagnetic spectrum do absorption bands occur that are diagnostic for different mineral types?
3. Under which conditions can signal mixing be considered a linear process?
4. Assume you will design a hyperspectral scanner sensitive to radiation in the 400 to 900 nm region. The instrument will carry 288 channels with a spectral resolution (bandwidth) of 1.8 nm. Does this configuration result in spectral overlap between the bands?

Chapter 14

Remote sensing below the ground surface

14.1 Introduction

When it comes to exploration for the resources of the solid earth, methods of detecting – directly or indirectly – resources of minerals, hydrocarbons and groundwater are needed that can "see" deep into the Earth. Similar capabilities may be required at a more local scale for foundation studies or environmental pollution problems. Methods that map the Earth's surface in or near the visible spectrum are a good start, and the interpretation of surface geology using RS images allows inference of what may lie below. However, there are other methods that actually probe more deeply into the ground by making use of the physical or chemical properties of the buried rocks. These properties – or changes in properties from one rock-type to another – are detected by carrying out geophysical surveys with dedicated sensors such as gravimeters, magnetometers and seismometers. Interpretation of geophysical surveys, along with all other available data, is the basis for planning of further investment in exploration, such as drilling, or using more expensive, dedicated geophysical techniques to probe in more detail into the most promising locations. Applied geophysical techniques form an important part of practical earth science, and the theory and application of the available methods is set out in several introductory textbooks ([20], [11], [6], [29]).

14.2 Gamma ray surveys

Thorium
Uranium
Potassium

Gamma radiation (electromagnetic radiation of very short wavelength; see Section 2.2) 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, gamma radiation mapping is limited to the shallowest subsurface. However, where soils are derived directly from the underlying bedrock and where bedrock outcrops exist, 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, therefore, not only counts the number of incoming rays (counts per second) but also, through analyzing the energy spectrum of all incoming gamma rays, attributes gamma-rays to their source elements and so estimates 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 14.1).

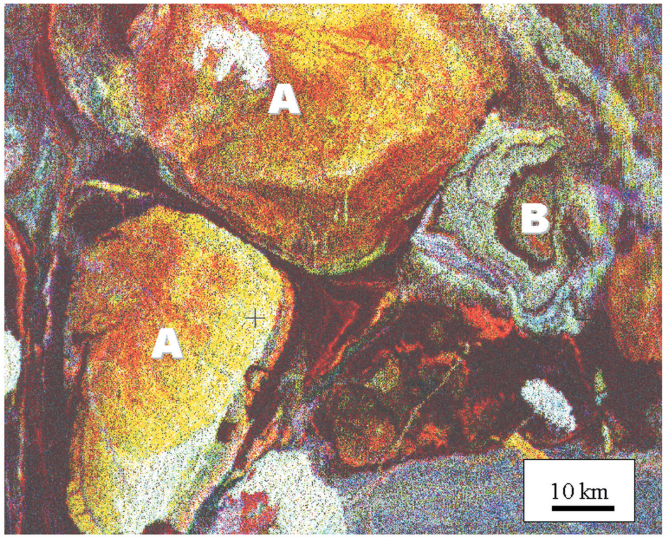


Figure 14.1: Ternary image (K=red, Th=green, U=blue) of an area of Archean geology in NW Australia. Domes of gneisses (A) show up as bright and red-orange in colour largely on account of their potassium content. The layers in an old sedimentary basin or syncline, cut through horizontally by erosion, are visible around B, where each layer has a different hue. Courtesy of Geoscience Australia.

14.3 Gravity and magnetic anomaly mapping

Gravity anomalies

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. 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 milligal (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

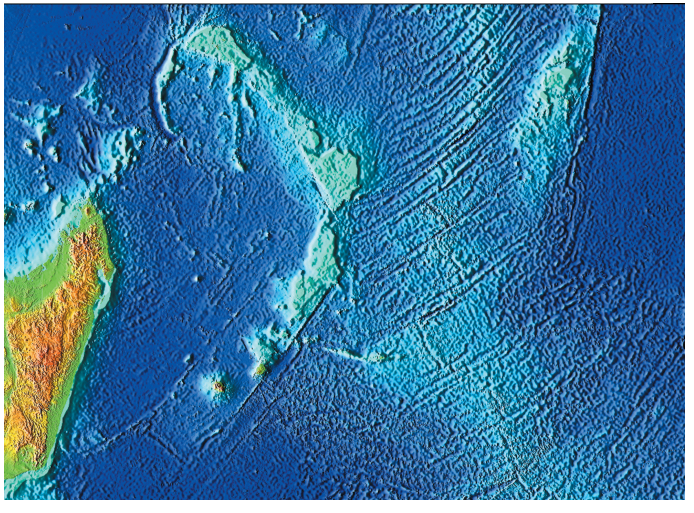


Figure 14.2: Sea floor relief in the western Indian Ocean, revealed by satellite altimetry of the sea surface (GEOSAT). Note the mid-ocean ridges and the transform faults either side of them. Note also the (largely sub-marine) chains of volcanic islands between India and Madagascar. Scene is roughly 3800 km across (from [26]).

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 elevation attributable to gravity anomalies. These arise from density variations in the subsurface, which, at sea, are due mainly to the elevation variations of the ocean floor. Mapping sea-level undulations has therefore made possible the mapping of sea-floor relief at the scale of a 5 km cell for all the world's oceans (Figure 14.2). In 2002 the *Gravity Recovery and Climate Experiment* (GRACE) satellite was launched. Built jointly by NASA and the German Aerospace Centre (DLR), GRACE is actually a tandem of two satellites equipped with GPS and a microwave ranging system, that allow an efficient and cost-effective way to map gravity globally with high accuracy.

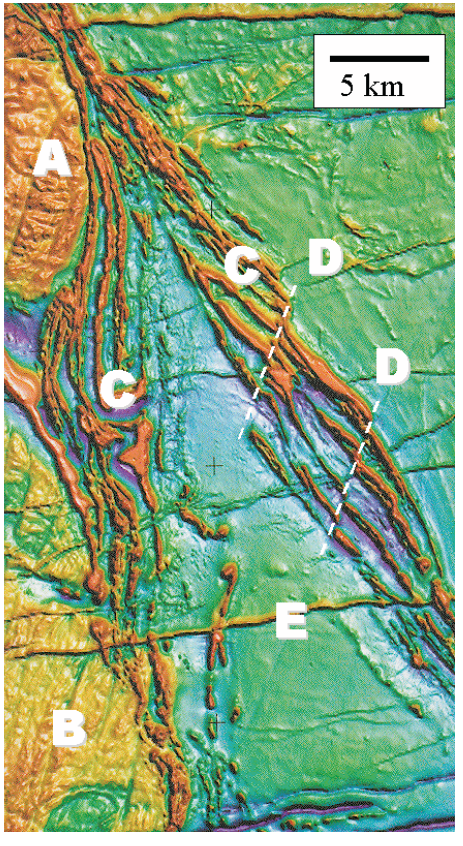
Gravity surveys

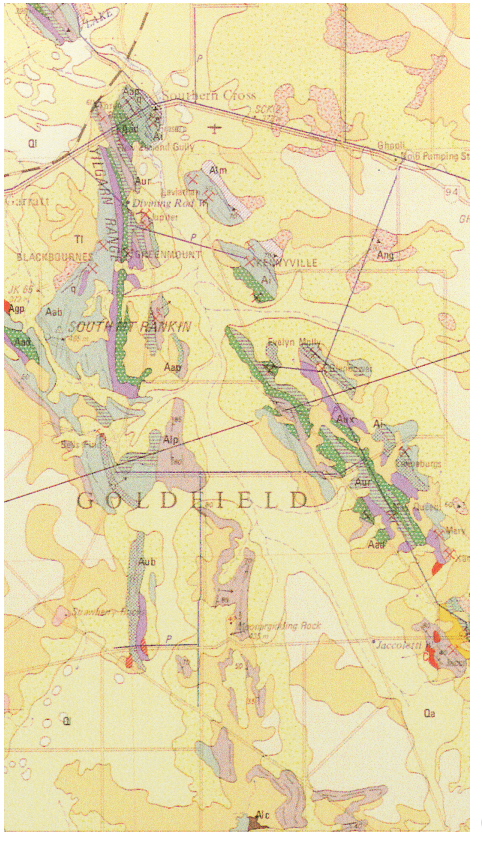
Mapping of magnetic anomalies from low-flying aircraft has been widely used in commercial exploration for over 50 years. The Earth's main field has a value that varies between 20,000 and 80,000 nano-Teslas (nT) over the Earth's surface. At a ground clearance of 50 to 100 metres, magnetic anomalies due 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 ppm of the total field. Each reading takes only 0.1 second, 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 (see Figure 14.3). 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.

Magnetic anomalies

Aeromagnetic surveys

(a)





(b)

Figure 14.3: (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 (*eg*, E) cut all formations. Note that faults such as D predate the emplacement of the dykes. Courtesy of Fugro Airborne Surveys; (b) Conventional geological map of the same area. Courtesy of Geoscience Australia and GSWA.

14.4 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.

Electrical methods

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 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 (see Figure 14.4). Other important ore bodies

EM methods

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.

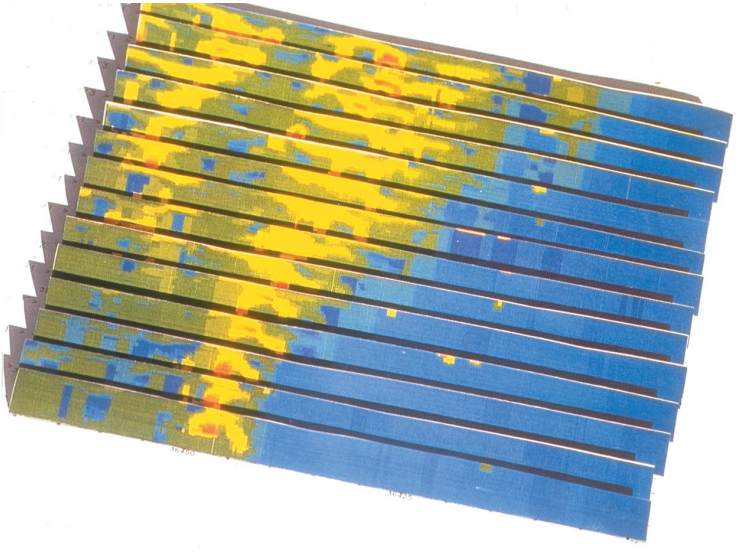


Figure 14.4: Conductivity cross-sections through the Bushman copper sulphide ore body in Botswana derived from airborne EM traverses flown east-west across the strike of the body. Red and yellow colours = highly conductive zones, blue = non-conductive zones. Courtesy of Fugro Airborne Surveys.

14.5 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.

Sub-audio

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 subsurface 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 (see Figure 14.5) 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

3D surveys

modest scales of exploration, can be applied for shallow investigations (depths of a few tens of metres), useful in groundwater exploration and site investigation.

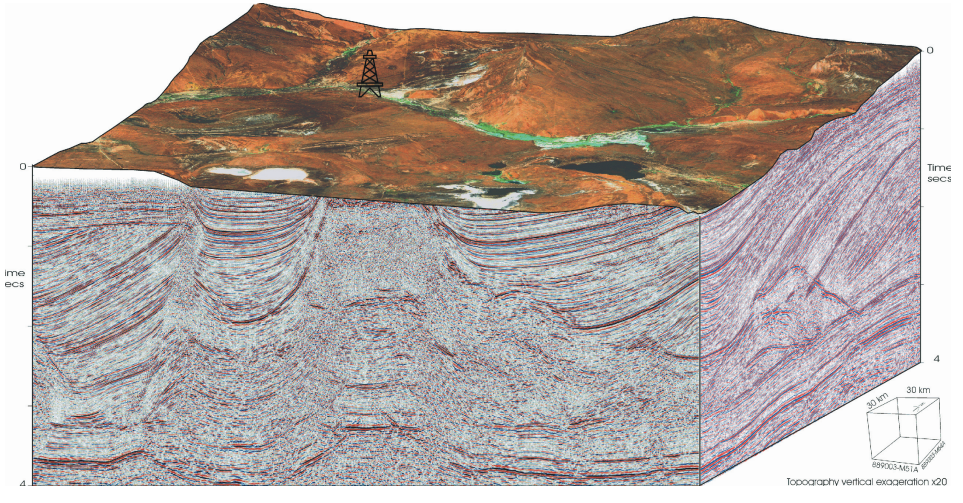


Figure 14.5: 3D seismic surveys map the layering in the subsurface, vital in oil exploration. The top surface of the cube is a RS image draped on a DTM (20 times vertical exaggeration). Side faces show seismic sections through the underlying strata.

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 over 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 photogeology, 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 often 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.

New technology is emerging to map and quantify the presence of water in the ground directly using nuclear magnetic resonance. If protons, such as those present in water, are aligned by applying a magnetic field, they precess when the applied field is turned off. Detection of the precession signal is, therefore, a direct indication of the presence of water below the surface. Magnetic resonance sounding and mapping promises to be a powerful tool for efficient groundwater exploration in the future.

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.

Ground penetrating radar: in many circumstances, radar signals of appropriate wavelength may be used to probe several metres - or even tens of metres - into the ground. The profiling and mapping of features such as the bedrock surface, buried pipelines and cables and old building foundations is useful in the context of highway construction, dam-site investigations and site engineering at

a scale that is very detailed in comparison to many geophysical investigations. Cost implies that systematic application of such detailed methods to larger areas is seldom an affordable option. Nevertheless, geophysical methods have successfully been brought to bear on a range of local problems such as archeological sites, unexploded ordnance, shipwrecks and pollution detection. In most of these cases, methods that do not probe beyond the ground (or water) surface are of comparatively little value.

Questions

The following questions can help you to study Chapter 14.

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 14.3(a) on a transparent overlay and compare your result with the geological map in Figure 14.3(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?

Bibliography

[1] A. Berk, L.S. Bernstein, and D.C. Robertson. *MODTRAN: A Moderate Resolution Model for LOWTRAN*. Air Force Geophysics Laboratory, Hanscom AFB, MA, US, 1997. 430

[2] W. Bijker. *Radar for Rain Forest: A Monitoring System for Land Cover Change in the Colombian Amazon*. Phd thesis, ITC, 1997. 373

[3] A. Brown A. and W. Feringa. *Colour Basics for GIS Users*. Prentice Hall of Canada Ltd, 2002. 215

[4] P.S. Chavez, S.C. Sides, and J.A. Anderson. Comparison of three different methods to merge multiresolution and multispectral data: Landsat tm and spot panchromatic. *Photogrammetric Engineering & Remote Sensing*, 57(3):295–303, 1991. 208

[5] R.N. Clark. Spectroscopy of rocks, and minerals,, and principles of spectroscopy. In A.N. Rencz, editor, *Manual of Remote Sensing: Remote Sensing*

for the Earth Sciences, volume 3, pages 3–58, New York, 1999. John Wiley & Sons. 467, 486

[6] M. Dobrin and C. Savit. *Introduction to Geophysical Prospecting*. McGraw-Hill, New York, US, forth edition, 1988. 489

[7] Elachi. *Spaceborne radar remote sensing: application and techniques*. IEEE Press, 1988. 408

[8] S.M.E. Groten. Land ecology - and land use survey. ITC Lecture Notes RUS10, 1994. 270

[9] J.R. Harris, C. Bowie, A.N. Rencz, and D. Graham. Computer enhancement techniques for the integration of remotely sensed, geophysical, and thematic data for the geosciences. *Canadian Journal of Remote Sensing*, 20(3):210–221, 1994. 208

[10] G.R. Hunt. Remote sensing in geology. In B. Siegal and A. Gillespie, editors, *Electromagnetic Radiation: The Communication Link in Remote Sensing*, page 702, New York, 1980. John Wiley & Sons. 486

[11] P. Kearey, M. Brooks, and I. Hill. *An Introduction to Geophysical Exploration*. Blackwell Science, Oxford, UK, 2002. 489

[12] F.X. Kniezys, E.P. Shettle, L.W. Abreu, J.H. Chetwynd, G.P. Anderson, W.O. Gallery, J.E.A. Selby, and S.A. Clough. *User Guide to LOWTRAN 7*. Air Force Geophysics Laboratory, Hanscom AFB, MA, US, 1988. 430

[13] H.J. Kramer. *Observation of the Earth and its Environment: Survey of Mission and Sensors*. Springer Verlag, Berlin, Germany, fourth edition, 2002. 163

[14] K. et al. Kraus. *User Manual SCOP*. University of Technology Vienna, Austria, 1998. 399

[15] T.M. Lillesand, R.W. Kiefer, and J.W. Chipman. *Remote Sensing and Image Interpretation*. John Wiley & Sons, New York, NY, fifth edition, 2004. 90, 107, 215, 222, 253, 277, 278, 313, 341, 432, 567

[16] K.R. McCloy. *Resource Management Information Systems*. Taylor & Francis, London, U.K., 1995. 83, 84, 341

[17] J.Ch. et al. McGlone. *Manual of Photogrammetry*. ASPRS, USA, 2004. 163, 253, 341

[18] H. Middelkoop. Uncertainty in a GIS, a test for quantifying interpretation output. *ITC Journal*, 1990(3):225–232, 1990. 274, 275

[19] K. Navulur. *Multispectral Image Analysis Using The Object Oriented Paradigm*. Taylor & Francis, 2007. 313

[20] D.S. Parasnis. *Principles of Applied Geophysics*. Kluwer Academic Publishing, Dordrecht, The Netherlands, 1996. 489

[21] N. Pfeifer and Ch. Ch. Briese. *Geometrical aspects of airborne laser scanning and terrestrial laser scanning*, volume XXXVI, Part 3/W52. IAPRS, 2007. 408

[22] W.R. Philipson. *Manual of Photographic Interpretation*, volume Second Edition. 2^{nd} edition. ASPRS, 1997. 277, 278

[23] H. Rahman and G. Dedieu. Smac: A simplified method for the atmospheric correction of satellite measurements in the solar spectrum. *International Journal of Remote Sensing*, 15:123–143, 1994. 433, 436

[24] R. Richter. *A Spatially-Adaptive Fast Atmospheric Correction Algorithm. ERDAS Imagine — ATCOR2 User Manual (Version 1.0)*, 1996. 432

[25] M. Schetselaar. On preserving spectral balance in image fusion and its advantages for geological image interpretation. *Photogrammetric Engineering & Remote Sensing*, 67(8):925–934, 2001. 210

[26] W. Smith and D. Sandwell. Measured and estimated seafloor topography, version 4.2. Poster RP–1, 1997. World Data Centre for Marine Geology and Geophysics. 493

[27] B. Sweetman. Jane's space directory 2006/2007. Technical report, Jane's Information Group, Alexandria, USA. 163

[28] D. Tanre, C. Deroo, P. Duhaut, M. Herman, J.J. Morcette, J. Perbos, and P.Y. Deschamps. Description of a computer code to simulate the satellite signal in the solar spectrum: the 5s code. *International Journal of Remote Sensing*, 11:659–668, 1990. 430

[29] W.M. Telford, L.P. Geldart, and R.E. Sheriff. *Applied Geophysics*. Cambridge University Press, Cambridge, UK, second edition, 1991. 489

[30] K. Tempfli. Interpolation and filtering. ITC Lecture Notes PHM102, 1982-1997. 238

[31] J.W. Trevett. *Imaging Radar for Resources Surveys*. Chapman and Hall Ltd., London, U.K., 1986. 372

[32] F. van der Meer and S. De Jong. *Imaging Spectrometry: Basic Principles and Prospective Applications*. Kluwer Academic Publishers, Dordrecht, the Netherlands, 2001. 465, 486

[33] E.F. Vermote, D. Tanre, J.L. Deuze, M. Herman, and J.J. Morcette. Second simulation of the satellite signal in the solar spectrum, 6s: an overview. *IEEE Transactions on Geoscience and Remote Sensing*, 35(3):675–686, May 1997. 430

[34] W.S. Warner, R. Graham, and R.E. Read. *Small Format Aerial Photography*. Whittles Publishing, Caithness, Scotland, 1995. 341

[35] A. Wehr and U. Lohr. Airborne laser scanning — an introduction and overview. *ISPRS Journal of Photogrammetry & Remote Sensing*, 54:68–82, 1999. 393

[36] I.S. Zonneveld, H.A.M.J. van Gils, and D.C.P. Thalen. Aspects of the ITC approach to vegetation survey. *Documents Phytosociologiques*, IV, 1979. 270

Glossary

Abbreviations & Foreign words

2D Two-dimensional.

3D Three-dimensional.

AARS Asian Association of Remote Sensing.

AATSR advanced along track scanning radiometer.

AC atmospheric correction.

AHN algemeen hoogte model.

AIS airborne imaging spectrometer.

ALI advanced land imager.

ALS airborne laser scanning.

ASAR advanced synthetic aperture radar.

ASTER advanced spaceborne thermal emission and reflection radiometer.

AVHRR advanced very high resolution radiometer.

AVIRIS airborne visibleinfrared imaging spectrometer.

B&W black-and-white (film, photographs).

BRDF bidirectional reflectance distribution function.

CAD computer aided design.

CCD charge-coupled device.

CHRIS compact high resolution imaging spectrometer.

CMOS complementary metal oxide semiconductor.

DEM digital elevation model.

DINSAR Differential InSAR.

DIY do it yourself.

DLR German Aerospace Centre (Deutsche Luft- und Raumfahrt).

DMC disaster management constellation.

DN digital number.

DPI dot per inch.

DSM digital surface model.

DTM digital terrain (relief) model.

EM electromagnetic.

EO earth observation.

EO-1 earth observing-1.

EOS Earth observing system.

ERS-1 (and ERS-2) European remote-sensing (satellites).

ESA European Space Agency.

FAO Food and Agriculture Organization.

FLI fluorescence line imager.

FOV field of view.

GCP ground control point.

GDA geospatial data acquisition.

GIS geographic information system.

GPS global positioning system.

GRC ground resolution cell.

GSD ground sampling distance.

HALE high altitude long endurance.

HRC high-resolution camera.

HRG (SPOT-5) high resolution geometric.

HRS (SPOT-5) high resolution stereoscopic.

ICESat ice, cloud, and land elevation satellite.

ICT information and communication technology.

IFOV instantanous field of view.

IGI Ingenieur-Gesellschaft für Interfaces.

IHS intensity-hue-saturation.

IMU inertial measuring unit.

INS inertial navigation system.

InSAR interferometric SAR.

IP induced polarization.

IR infrared.

IRS Indian remote sensing.

ISPRS International Society of Photogrammetry and Remote Sensing.

ISRO Indian Space Research Organisation.

ISS international space station.

ITC International Training/Torture/Tourist Centre.

LAC LEISA atmospheric corrector.

LEISA linear etalon imaging spectrometer array.

LIDAR light detection and ranging.

LISS linear imaging self-scanning sensor.

LUT look-up table.

MDM minimum distance to mean.

MERIS medium resolution imaging spectrometer.

MIT Massachusetts Institute of Technology.

ML maximum likelihood.

MODIS moderate-resolution imaging spectroradiometer.

MSG meteosat second generation.

MSS MultiSpectral Scanner (Landsat).

NASA National Aeronautics and Space Administration.

NDVI normalized difference vegetation index.

NIR near-infrared.

NOAA National Oceanic and Atmospheric Administration.

OOA object-oriented analysis.

OSA optical sensor assembly.

PAN panchromatic.

PC personal computer.

PCA principal component analysis.

PCC proportion correctly classified.

PMI programmable multispectral imager.

PoRS principles of remote sensing.

POS positioning and orientation system.

ppm part per million.

PRI primary data (used in "pri format" of image files).

Proba project for on-board autonomy.

RADAR radio detection and ranging.

RAR real aperture radar.

RGB red-green-blue.

RMSE root mean square error.

RPC rational polynomial coefficients.

RS remote sensing.

RTM radiative transfer model.

SAM spectral angle mapper.

SAR synthetic aperture radar.

SEVIRI spinning enhanced visible and ir imager.

SI international system of units (Le Système International d'Unités).

SIS scanning imaging spectroradiometer.

SLC single-look-complex.

SMIRR shuttle multispectral infrared radiometer.

SONAR sound navigation ranging.

SPOT Satellite Probatoire pour l'Observation de la Terre.

SRTM shuttle radar topography mission.

SWIR short-wave infrared.

TES technology experiment satellite.

TIR thermal infrared.

TLS terrestrial laser scanning.

TM/ETM thematic mapper / enhanced thematic mapper.

TOA top of the atmosphere.

UAV unmanned aerial vehicle.

USGS USA Geological Survey.

UTM Universal Transverse Mercator (map projection).

UV ultraviolet.

VIS visible (bands).

VNIR visible and near-infrared (bands).

WiFS wide field sensor.

WMO World Meteorological Organization.

YMC yellow, magenta, cyan.

Terms

A/D conversion Acronym for analogue to digital conversion; the process of sampling an analogue signal and quantifying signal strength by a number (stored as binary code).

Absorption The process in which electromagnetic energy is converted in an object or medium into other forms of energy (*eg*, heat, or fluorescence) or causing chemical reactions (*eg*, photosynthesis).

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, *eg*, on computer screens and television sets.

Aerial camera A camera specially designed for use in an aircraft (opposed to a terrestrial camera or spaceborne camera). In the context of this book mostly used to denote an aerial survey camera. A **survey camera** is specially designed to produce images for the purpose of surveying and mapping (also referred to as metric camera; can be aerial or terrestrial). An aerial survey camera is typically designed for being mounted in an airplane to take vertical photos/images of large format; it can be a film camera or a digital frame camera or line camera. A **frame camera** is a camera in which an entire frame or format is exposed simultaneously through a lens (with a fixed focal length).

Aerial triangulation The process of finding the exterior orientation parameters of a block of aerial photos/images using a (very) limited number of ground control points.

Aerosol A suspension of fine solid (dust) or liquid particles (water vapour) in a gas (air).

Affine transformation A 2D geometric transformation (plane-to-plane transformation) that uses six parameters to account for rotation, translation, scale change, and shearing. It defines the (linear) relationship between two coordinate systems such as an image coordinate system and a map coordinate system; see also **conformal transformation**.

Algorithm A procedure for solving a mathematical problem (as of finding the greatest common divisor) in a finite number of steps.

Altitude The elevation of a (moving) object above a reference surface, usually mean sea level.

Amplitude In the context of this book, the maximum departure of a wave from its average value; it is a measure for the strength of an oscillatory movement. For instance, the louder a sound, the larger the amplitude of the sound wave.

Aperture In optics, an aperture is an opening through which light is admitted; it determines how collimated the admitted rays are. For a conventional camera, the aperture is materialized by the diaphragm as part of the lens. In radar, the aperture is related to the antenna length.

Atmosphere The Earth's atmosphere is the gaseous envelope surrounding the Earth. The atmosphere is classified into several layers, but there are no discernible boundaries; it gradually becomes thinner. Three quarters of the atmosphere's mass is within 11 km of the Earth's surface.

An altitude of 120 km marks the boundary where atmospheric effects become noticeable during re-entry from space.

Atmospheric scattering The process of particles or gaseous molecules present in the atmosphere redirecting EM radiation from its original path. A well known example is Rayleigh scattering.

Atmospheric window A spectrum portion outside the main absorption bands of the atmospheric gases that can be used for remote sensing.

Azimuth In mapping and navigation azimuth is the direction to a target with respect to north and usually expressed in degrees. In radar RS azimuth pertains to the direction of the orbit or flight path of the radar platform.

Backscatter The microwave signal reflected by elements of an illuminated surface in the direction of the radar antenna.

Band In the context of this book, mostly short for 'wavelength band', which stands for a limited range of the EM spectrum. A sensor is sensitive to certain 'spectral bands'; see also **spectral band**. Atmospheric absorption is characterized by 'absorption bands'. The term 'band' is also frequently used to indicate one of the digital images of a multi-band image, thus the data recorded by one of the channel of a multispectral sensor; *eg*, band 3 of a Landsat image, or the green band of a SPOT image.

Black-body A body/object that absorbs all EM energy that hits it.

Blunder A gross error resulting from a misrecording or incorrect data entry, incorrect measurement of a ground control point in an image, etc.

Brightness Luminance in photometric terms: the luminous intensity of a surface in a given direction per unit of projected area (equivalent to radiance in radiometry, but expressed in candela per square meter). In the book loosely used for the degree of illumination or the state of being bright; very bright seen as white, very dark seen as black. In the section on radar, 'brightness' is used as the property of a radar image in which the observed strength of the radar reflectivity is expressed as being proportional to a digital number.

Bundle of rays A photogrammetric term to denote the geometric construct that is obtained by connecting each point measured in an image with the perspective centre by a straight line (the imaging light ray). There is one bundle of rays for each photograph and it is used for exterior orientation or aerial triangulation.

Calibration report In aerial photography, the manufacturer of the camera specifies the interior orientation in the form of a certificate or report. Information includes the principal distance, principal point, radial lens distortion data, and fiducial mark coordinates in the case of a film camera.

Camera An electro-optical remote sensor without mechanical components. In its simplest form, it consists of the camera body, a lens, a focal plane array of CCDs, and a storage device. We distinguish line cameras and frame cameras, depending on whether they have (a few) single line(s)

of photosensitive cells or a matrix array of cells (CCDs or CMOSs); see also **aerial camera**.

CCD (Charge-Coupled Device) A solid state detector using silicon as semi-conductor material. Many are assembled on one chip and equipped with a shift register to transport charge and sample it at the end of a CCD line after amplification. One CCD is good for one pixel in one spectral band. A camera with three channels needs three CCDs for one ground resolution cell. CCDs are used in cameras for sensing in spectral bands with wavelengths smaller than about 1 μm. Solid state detectors that are sensitive to SWIR and even TIR radiation use other semi-conductor material and they need IR blockers to reduce noise and thus increase the dynamic range.

Channel In the context of this book, the sensor component for a particular spectral band (detector, electrical circuit, storage). Many remote sensors also have one or more communication channels.

Check point An additional ground point used to independently verify the degree of accuracy of a geometric transformation (*eg*, georeferencing, exterior orientation, aerial triangulation, orthophoto).

Class A group, set, or kind sharing common attributes. In image interpretation and classification, we define thematic classes of terrain features; they are variables of the nominal type. For instance, typical classes of land cover are grass, wheat, trees, bare soil, buildings, etc.

Cluster Used in the context of image classification to indicate a concentration of observations (points in the feature space) related to a training class.

CMOS (Complementary Metal Oxide Semiconductor) A device similar to a CCD. It also has a silicon photodiode, however, a CMOS array does not have a shift register but each cell has its own MOS switch for readout.

Collinearity A (nonlinear) mathematical model that photogrammetric restitution is based upon. **Collinearity equations** describe the relationship among image/camera coordinates, object space coordinates, and orientation parameters. The assumption is that the exposure station of a photograph, a ground point, and its corresponding image point location all lie along a straight line.

Colour A sensation of our visual system caused by EM radiation; a visual perception property; a phenomenon of light. Different colours are caused by different wavelengths of light. We may quantify 'colour' by hue (in the IHS colour 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 Photographic film with specific sensitivity for near-infrared wavelengths. Typically used in surveys of vegetation.

Cones Photoreceptor cells in the retina of the eye for sensing colour; see also **rods**.

Conformal transformation A 2D geometric transformation (plane-to-plane transformation) that uses four parameters to account for rotation, trans-

lation, and scale change. It defines the (linear) relationship between two coordinate systems such as an image coordinate system and a map coordinate system; see also **affine transformation**.

Coordinates Linear or angular quantities, which designate the position that a point occupies in a given reference frame or system. The **coordinate system** is based on mathematical rules and used to measure distances and/or angles in order to identify the location of points by means of unique sets of numerical values. A coordinate system requires the definition of its origin, the orientation of the axes, and units of measurement. Plane rectangular coordinates describe the position of points with respect to a defined origin by means of two distances perpendicular to each other; the two reference lines at right angles to each other and passing trough the origin are called the **coordinate axes**.

Corner reflection In radar RS: a high backscatter as typically caused by two flat surfaces intersecting at 90 degrees and situated orthogonally to the radar incident beam.

Corner reflector The combination of two or more intersecting specular surfaces that combine to enhance the signal reflected back in the direction of the radar, *eg*, houses in urban areas.

D/A conversion Acronym for digital to analogue conversion; the process of converting a binary code to a charge (an electrical signal).

Data Factual information in digital form, used as basis for calculation, reasoning by a computer, etc.

Detector A device or medium for detecting the presence of (electromagnetic) radiation.

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) A representation of a surface in terms of elevation values that change with position. Elevation can refer to the ground surface, a soil layer, etc. According to the original definition data should be in a raster format.

Digital number (DN) The recorded digital read-out of an electronic detector. It is the quantized sampled value of the electrical signal which is generated by the detector. The DNs correspond to photon energy incident upon the detector and radiances at the detector, but have not a meaningful physical unit. In 8 bits recording, the DNs are in the range [0, 255].

Digital surface model (DSM) In earth observation from air or space, often synonym for DEM, elevation standing for terrain elevation.

Digital terrain model (DTM) A digital representation of terrain relief in terms of (X,Y,Z) coordinates and possibly additional information (on breaklines and salient points). Z usually stands for elevation, and (X,Y) for the horizontal position of a point. To the concept of a DTM it does not matter whether Z is orthometric or ellipsoidal elevation. Horizontal position can be defined by geographic coordinates or by

grid coordinates in a map projection. DTM data can be given in different forms (controur lines, raster, TIN, profiles, etc).

Digitizing Generally, any process that converts an analogue representation of a physical quantity to a digital representation, *ie*, numeric values. Specifically, in the context of this book, the process of manually tracing lines or other image features to obtain coordinates of these features (a 'vector representation'). We can either digitize features (a) on hardcopy images using a digitizing tablet (2D) or analogue/analytical photogrammetric plotter (3D), or (b) on softcopy images using a standard computer screen and mouse (2D **on-screen digitizing**) or a digital photogrammetric workstation (3D). The process of converting an entire hardcopy image to a digital image (*ie*, a raster representation) is referred to as scanning rather than digitizing.

Dynamic range The ratio between the maximum level and the minimum level of intensity that can be measured, also called the signal to noise ratio of the detector; it is commonly expressed in decibels (db).

Earth observation The process of gathering information about physical, chemical, biological, geometrical properties of our planet.

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 range of all wavelengths, from gamma rays (10^{-12} m) up to very long radio waves (10^{8} m).

Elevation The height of a surface (*eg*, the terrain surface) above a reference surface (*eg*, an ellipsoid or the geoid); see also **height** and **altitude**.

Emission The act or instance of releasing (specifically EM) energy from a body.

Emissivity of a material is the ratio of energy emitted by a particular material to energy radiated by a black-body at the same temperature. It is a measure of a material's ability to absorb and re-emit energy. Emissivity is a dimensionless quantity.

Emittance of a body/object quantifies how much energy is emitted by it. Instead of 'emittance' sometimes the term **exitance** is used. **Spectral (radiant) emittance** is emittance per wavelength (quantifying the spectral distribution of emitted energy); it is measured in $Wm^{-2}m^{-1}$.

Energy A fundamental entity of nature, which exists in many forms including kinetic, potential, chemical, acoustic, electromagnetic energy, etc. Radiant energy is the energy of electromagnetic waves or conceptualized as the energy carried by a stream of photons. The term **radiant energy** is most commonly used in the fields of radiometry, heating and lighting.

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.

Exposure station A term of aerial photography: the point in the flight path at which the camera exposes the film or opens the shutter for expos-

ing a CDD frame. The exposure station has elements that define the position of the projection centre and the camera attitude.

Exterior orientation See **orientation**.

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.

Features A set of measureable properties, *eg*, of the Earth's surface. *Eg*, we may call the set of objects resulting from human construction work - the houses, roads, irrigation channels - cultural terrain features.

Fiducial marks Four or eight reference markers fixed on the frame of a film survey camera and visible in each photograph. Fiducials are used to compute the transformation from pixel coordinates to image/camera coordinates.

Field of view (FOV) The viewing range of a sensor expressed as angle, usually in degrees; sometimes referred to as swath angle, or in ALS as scan angle. The FOV is one of the factors which determine the swath width of a sensor-platform system; see also **instantaneous field of view**.

Filter (a) Physical device to suppress an input component; *eg*, a yellow filter in front of a lens of a camera (it absorbs blue light). (b) Algorithm in

signal/image processing for eliminating or at least reducing an unwanted component of given data, *eg*, a noise filter; see also **filtering**.

Filtering Computational process of changing given values such that a contained component is either attenuated, amplified, or extracted (*eg*, smoothing an image or sharpening an image, extracting brightness edges in an image, removing off-ground points from a DSM to obtain a DTM, etc).

Flying height The vertical distance between the camera/sensor position at the time of exposure and the terrain at average elevation within the ground coverage of the photo/image taken.

Focal length The distance between the optical centre of the lens and where the optical axis intersects the plane of critical focus of a very distant object. The focal length of a survey camera is determined in a laboratory; see also **principal distance**.

Focal plane The plane (perpendicular to the axis of the lens) in which images of points in the object space of the camera are focused.

Foreshortening Spatial distortion whereby terrain slopes facing the side-looking radar are mapped as having a compressed range scale relative to its appearance if the same terrain was flat.

Frequency The reciprocal of the wave period.

Gain of a linear filter or kernel, see **kernel**.

Geocoding The 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; see also **georeferencing**.

Georeferencing The 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 is an image and coordinates of ground control points, output are the transformation parameters (a "georeferenced image").

Geospatial data Factual information related to location on the (surface of the) Earth.

GPS A satellite surveying method providing accurate geodetic coordinates for any point on the Earth at any time.

Grey values Grey is a mixture of black and white. The computer converts the DNs (in the range 0 to 255) of a digital image to grey values on the monitor; 0 becomes black 255 white.

Grey-body A body or material that absorbs and radiates only a certain fraction of EM energy compared to a black-body.

Grid A network of regularly spaced horizontal and perpendicular lines (as for locating points on a map). We may associate (field) values with the nodes of a grid, *eg*, elevation to obtain a DEM; see also **raster**.

Ground control points (GCP) A ground point reliably identifiable in the image(s) under consideration. It has known coordinates in a map or terrain coordinate system, expressed in the units (*eg*, metres, feet) of the specified coordinate system. GCPs are used for georeferencing and image orientation.

Ground range Range between the radar antenna and an object as given by a side-looking radar image but projected onto the horizontal reference plane of the object space.

Ground resolution cell (GRC) The area on the ground corresponding to the IFOV of a detector. The area can be elliptical or rectangular. A pixel value (DN) corresponds to the averaged radiance of the ground resolution cell. The extent of the GRC is sometimes referred to as pixel size (on the ground).

Ground sampling distance (GSD) The distance between the centres of two adjacent resolution cells of the same sensor channel. For scanners and line cameras the GSD can be different along track and across track. The GSD equals the extent of a ground resolution cell in a well designed imaging sensor. The GSD is also referred to as pixel size (on the ground).

Ground surface The bare Earth's surface (also referred to as "bald Earth").

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, whereas this may be difficult to achieve.

Heat The quality of being hot or the condition of matter which produces the sensation warmth. It is one of the primary sensations, produced by contact with or nearness to fire or any body at a high temperature.

Height The vertical extent of an object, the distance from the bottom to the top of something protruding above a level.

Histogram Tabular or graphical representation showing the (absolute or relative) frequency of values of a variable. In the context of digital images it relates to the distribution of the DNs of a set of pixels.

Histogram equalization The process used in the visualization of digital images 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.

Hue Quantification (in the IHS colour space) of what we refer to as blue, green, yellow, orange, red purple, etc.

Image In the context of this book, the optical counterpart (pictorial representation) of an object produced by an optical device or an electronic device. An example of an image is a photograph, which is the likeliness of an object or scene recorded on photographic material. Another example is the picture produced on a computer screen or a television set. The term "remote sensing image" is frequently used to either distinguish arbitrary images on an electronic display from those originating from a sensor or to denote raw data produced by an electronic sensor, which are in fact not pictorial but arrays of digital numbers; the digital numbers are related to a property of an object or scene,

such as the amount of reflected light. Similarly also the term **digital image** is commonly used for an array of digital numbers, which can readily be converted to an image on a computer screen or by a printer. It is convenient to call the result of scanning a photograph or the data produced by a digital camera 'digital images'.

Image classification The process of assigning pixels to nominal, *ie*, thematic, classes. Input is a multi-band image, output is a raster in which each cell has a (thematic) code. Image classification can be accomplished using a supervised or unsupervised approach.

Image coordinate system A system of expressing the position of a point in an image by plane rectangular coordinates. You will find different definitions of an image coordinate system in literature. In this book, we define it as either the row-column system of a digital image, or the x-y photo system ("frame image coordinates at the time of exposure") with the principal point as origin and the x-axis in the direction of flight or the respective fiducial mark.

Image enhancement The process of improving the visual representation of a digital image, *eg*, by a histogram operation or using filters.

Image interpretation The key process in information extraction from images. The application context determines what is to be considered as information. We can use visual interpretation or computer vision techniques for recognizing features and objects of interest in an image.

Image matching The process of matching features common to two or more images (finding conjugate points or lines). Digital image matching

is used for various purposes; main applications are automated DSM generation and automated image orientation.

Image processing system A computer system that is specifically designed to process digital images and to extract information from them by visualizing the data, or applying models and pattern recognition techniques.

Image sensor (or imager) A photographic camera is an imaging sensor. The term, however, is mostly used for (optical-)electronic sensors. They provide data of a scene in an image fashion in the form of a two-dimensional array of DNs for each spectral band of sensing. A single element of such a 2D array is referred to as pixel. The pixel value - the DN - is an integer number in a fixed range. The range is a power of 2, depending on how many bits are used for storing a DN. 8 bits is very common, but it can be up to 16 bits especially for thermal and microwave sensors. Such an array - the '**digital image**' - can readily be used to drive a computer monitor or a printer after D/A conversion, this way creating an image.

Image space The mathematical space describing the (relative) positions of the observations. Image positions are expressed by their row-column index.

Imagery is mostly used as high-sounding term for images; it is avoided in this book.

Incidence angle In the context of imaging radar, the angle between the line of sight from the sensor to an element of an imaged scene and a verti-

cal 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 ground resolution cell.

Inertial measuring unit (IMU) A device providing us with attitude data of a sensor it is attached to or of an aircraft in which it is located. An IMU is the core instrument of an INS and it is also used in a POS for direct sensor orientation.

Instantaneous field of view (IFOV) The viewing range of a detector expressed as angle, usually in milliradians; referred to in active sensing (ALS) as beam divergence of the radiation. The IFOV is one of the factors, which determine the size of the ground resolution cell of a sensor-platform system; see also **field of view**.

Instrument In the context of GDA, a measuring device for determining the present value of a quantity under observation.

Intensity The term is used in different disciplines with different meanings and quantifications. In this book - except for the section on colour - it is used in a common language sense to express an amount of energy, irrespective of the unit in which it is measured (radiance, irradiance, emittance, spectral emittance, etc). In radiometry, **radiant intensity** is the radiant power from a point source and measured in Wsr^{-1}. See also **radiance** for comparison.

Interference In radar RS: the wave interactions of the backscattered signals from the target surface.

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.

Interior orientation See **orientation**.

Interpolation (from Latin *interpolire*, putting in between) Estimating the value of a (continuous) variable that is given by n sampled values at an intermediate point or instant. *Eg*, temperature has been measured at every turn of the hour; computing a temperature value at 10:37 from the 24 values given for a day is a matter of interpolation.

Interpretation elements A set of cues used by the human vision system to interpret a picture. The seven interpretation elements are: tone/hue, texture, pattern, shape, size, height/elevation, and location/association.

Interpretation key A guideline for image interpretation, which relates terrain features to observable features of an image.

Interpretation legend A description of interpretation units in terms of interpretation elements.

Irradiance The amount of incident energy on a surface per unit area and per unit time. Irradiance is usually expressed in Wm^{-2}.

Kernel The *n* by *m* array of coefficients used by a linear filter (moving average) to compute an output pixel value from the given values at the

pixel under consideration and its neighbours. A kernel can only define a linear filter if the input (and the output) of filtering are raster data. The reciprocal of the sum of the kernel values is the 'gain' of the kernel.

Laser Acronym for Light Amplification by Stimulated Emission of Radiation. Laser instruments used as active sensors for ranging and imaging operate in the visible to near-IR range. Topographic **LIDAR** (LIght Detection And Ranging) instruments use wavelengths between 0.9 to 1.6 μm.

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 image is transformed into a visible image by the development process in which the exposed silver halide is converted into silver grains that appear black.

Latitude/Longitude (Lat/Lon) The coordinate components of a spherical coordinate system, referred to as **geographic coordinates**. The latitude is zero on the equator and increases towards the two poles to a maximum absolute value of 90°. The longitude is counted from the Greenwich meridian positively eastwards to the maximum of 180°.

Layover Extreme form of foreshortening, *ie*, relief distortion in a radar image, in which the top of the reflecting object (*eg*, a mountain) is closer to the radar than the lower part of the object. The image of such a feature appears to have fallen over towards the radar.

Least squares adjustment A method of correcting observations in which the

sum of squares of all the residuals derived by fitting the observations to a mathematical model is made a minimum. Least squares adjustment is based on probability theory and requires a (large) number of redundant measurements.

Light Electromagnetic radiation that is visible to the human eye; also referred to as radiant energy enabling visual perception. We see light in the form of colour. The wavelength range is 0.38 to 0.76 μm.

Look angle The angle of viewing relative to the vertical (nadir) as perceived from the sensor.

Map A simplified (purpose-specific) graphical representation of geographic phenomena, usually on a planar display. From an EO perspective, it is a conventionalized image of reality, with documented conventions on the type of abstraction, the symbolic presentation, and the mathematical projection of 3D reality (including scale). Similar to stretching the notion of image to digital image we also use the term digital map.

Map coordinate system A system of expressing the position of a point on the Earth's surface by plane rectangular coordinates using a particular map projection, such as UTM, the Lambert's conical projection, or an azimuthal stereographic projection (as used in the Netherlands).

Mapping In the context of GDA, the process of converting (RS) data or images to a conventionalized image, the map. Mapping can be merely a radiometric and/or geometric transformation of an image, or involve

information extraction by visual interpretation or automated classification.

Measurement An observation yielding a quantitative record. In 1896, G. Grigorou won the first marathon race; it took him 3 hours 45 minutes to do the 42.195 km.

Microwaves Electromagnetic radiation in the microwave window, which ranges from 1 to 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.

Monochromatic (Derived form Greek *monochrõmatos*, of a single colour). Monochromatic light is light of a single wavelength. The closes realization of monochromatic radiation in remote sensing is offered by a topographic laser rangefinder with $\lambda = 1550 \pm 5$nm (NIR laser) or a bathymetric laser rangefinder with $\lambda = 532 \pm 4$nm (green laser). Note that with this understanding there is no such thing as "monochrome film". You may come across this term in photography books for black-and-white (B&W) film. Neither 'black' nor 'white' are colours!

Monoplotting The process that enables extraction of accurate *(x,y)* coordinates from an image by correcting for image distortions, in particular relief displacement.

Multispectral scanning A remote sensing technique in which the Earth's surface is scanned and reflected radiation is recorded simultaneously in different wavelength bands.

Nadir The point/area on the ground vertically beneath the sensor at the moment of imaging a line or area of the scene, sometimes referred to as ground nadir. In the case of a camera, it is the vertical projection of the optical centre of the lens. The image of the ground nadir is referred to as **nadir point**, or sometimes as photograph nadir.

Noise Any unwanted or contaminating signal competing with the signal of interest.

Object An entity obtained by abstracting the real world, having a physical nature (certain composition of material), being given a descriptive name, and observable; *eg*, "house". An object is a self-contained part of a scene having certain discriminating properties.

Object space The three-dimensional region that encompasses the physical features imaged by a remote sensor; see also **image space**.

Observation An act of recognizing and noting a fact or occurrence often involving measurement with instruments (earth observation) as well as a record or description so obtained. The outcome can be qualitative or quantitative. My girl friend is pretty, the boy is handsome, the fish smells badly.

Optics The branch of physics that deals with the propagation of light and the interaction of light with matter. While originally restricted to light, RS has stretched 'optics' to include the description of behaviour of UV and infrared radiation because of its similarity to 'sight'. Optics explains behaviour such as reflection and refraction. **Optical instru-**

ments/devices use components such as lenses, glass prisms, and mirrors.

Orbit The path followed by one body (*eg*, a satellite) in its revolution about another (*eg*, the Earth).

Orientation In photogrammetry, the process of relating some form of image coordinate system to some form of 3D system. We distinguish interior, exterior, relative and absolute orientation. Interior orientation provides internal camera model information that describes the construction of the bundle of rays (as needed for exterior orientation) using the principal distance, principal point (and lens distortion). Exterior orientation provides external camera model information that describes the exact position and rotation of each image in an object space system as they existed when the image was taken. Relative orientation computes the mutual pose of two (stereoscopic) images at the time of exposure; subsequent absolute orientation establishes the relationship with an object space coordinate system.

Orthophoto/orthoimage An aerial photo or satellite image that has been transformed to a map projection and radiometrically enhanced such that it is free of (significant) geometric distortions and radiometric disturbances.

Overlap (forward) Considering a traditional aerial camera, when two images overlap, they share a common area. *Eg*, in a block or strip of photographs, adjacent/subsequent images typically overlap by 60% .

Panchromatic 'Sensitive to light of all colors'. For instance, the rods of our

eyes are panchromatic sensors: they are not sensitive to light of a specific colour but only distinguish intensity differences of light (brightness sensed over the entire visible range). Related terms are polychromatic (multicoloured) and monochromatic (single coloured).

Parallax The apparent displacement of a point (*eg*, as seen in a central perspective image) with respect to a point of reference or coordinate system, caused by the shift in the point of observation. The stereoscopic parallax - considering a stereo pair of photographs of equal principal distance - is the algebraic difference of the distances of the two images of the point from the respective nadir points of the photos measured parallel to the air base.

Pattern As an interpretation element, it refers to the spatial arrangement of features in an image; it implies the characteristic repetition of certain forms or relationships.

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.

Period The time interval of two successive maxima of an oscillation; the duration of one cycle of a wave.

Phase In the context of EM wave theory, the shift parameter for the starting point of a wave.

Photo block A block of aerial photographs is the set of images obtained by an aerial survey mission. A traditional frame camera block might consist of a number of parallel strips with a sidelap of 20- 30% and a forward overlap of 60%. Photogrammetric software typically stores all of the information associated with a photogrammetric mapping project in the **block file**, such as: spheroid and datum, map projection; camera model, ID of images, GCPs, orientation coefficients, etc.

Photogrammetry The science and technique of making measurements on photographs and converting these to quantities meaningful in the terrain. In **analog photogrammetry**, optical and mechanical instruments, such as analog plotters, are used to reconstruct 3D geometry from two overlapping photographs. In **analytical photogrammetry** the computer replaces some expensive optical and mechanical components and software relates image space(s) to object space. In **digital photogrammetry** (hardcopy) photographs are replaced by digital images.

Photograph An image on photographic material, film or paper. A photograph in its strict sense is analogue and a record of reflected electromagnetic energy of an object or scene of only a very narrow spectral range from ultraviolet to near infrared. In an even stricter sense, light sensitive film should also have been used by the sensor to detect the electromagnetic energy. (Note, according to this very strict definition a picture taken by a digital camera and printed on an ink-jet printer is not a photograph nor is a panchromatic image produced from raw data of an electronic camera on a photographic film writer). **Aerial photographs** are photographs taken from positions above the Earth by a camera on an aircraft.

Photography The process or art of producing images by light on a sensitive surface (and subsequent development the exposed film).

Photon The elementary energy carrier used to explain electromagnetic energy in particle theory; it travels at the speed of light and has both particle and wave properties.

Picture A (2D) counterpart of an object or scene produced by a device or a human (artist). A photograph is an image, an image is a picture, but not all pictures are images and not all images are photographs.

Pixel The term stands for 'picture element'; it is the building cell of a digital image; see also **imaging sensor**.

Pixel value The digital number (DN) a pixel takes.

Platform A vehicle, such as a satellite or aircraft (or part of it), used to carry a sensor.

Polarization In the context of this book, the action or state of affecting radiation and especially EM radiation so that the vibrations of the wave assume a chosen orientation. In radar RS, we may use a signal where the electric field oscillates vertically or a microwave with a horizontally oscillation electric field, thus obtaining different images. For stereoscopic vision, we can use horizontally respectively vertically polarized light to distinguish the left image from the right image when overlaid on a computer monitor and viewed with special spectacles (one glass allowing vertically polarized light to pass through, while the other allows only horizontally polarized light).

Polychromatic 'Comprising many colours'. Solar radiation is polychromatic; see also **monochromatic**.

Principal distance The perpendicular distance from the perspective centre to the plane of an image of a digital camera or a particular finished photograph. The distance is equal to the calibrated focal length (but in the case of a film camera, corrected for the film or paper shrinkage or enlargement).

Principal point The intersection point of the perpendicular from the projection centre of a camera with the image plane.

Projective transformation A 2D geometric transformation that uses eight parameters to mathematically model perspective imaging of a plane. It defines the relationship between two coordinate systems such as the coordinate system of a central perspective image and a map coordinate system or a local coordinate system of a plane in object space; see also **conformal** and **affine transformation**.

Pulse An EM wave with a distribution confined to a short interval of time. Such a distribution is described in the time domain by its width and its amplitude.

Quantization The number of discrete levels applied to store the energy as measured by a sensor, *eg*, 8 bits quantization allows 256 levels of energy.

Radar Acronym for Radio Detection And Ranging. Radar instruments are active sensors, sensing at wavelengths between 1 and 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.

Radian The SI unit of angle in a plane, equal to $180/\pi$ degrees.

Radiance The amount of energy being emitted or reflected from a particular area per unit solid angle and per unit time. Radiance (observed intensity) is in photometry the radiant power from an extended area and measured in $Wm^{-2}sr^{-1}$; (**power** is also called **flux** - it is radiant energy per unit time, measured in watt). **Spectral radiance** is radiance per wavelength (band).

Radiant energy The energy of electromagnetic waves; the term radiant energy is most commonly used in the fields of radiometry, heating and lighting.

Radiometer A sensor, which measures radiant energy and typically in one broad spectral band ('single-band radiometer') or in only a few bands ('multi-band radiometer'), but with high radiometric resolution.

Radiometric resolution The degree to which intensity levels of incident radiation are differentiated by a sensor; usually expressed as the number of bits used for storing a DN. The number of bits used defines the quantization increment in A/D conversion. In one bit recording a radiance becomes either 0 or 1. We can show DNs on a computer monitor as grey values. If the raw data represent a binary image, then 0 is displayed as black and 1 as white. In the case of 8 bits recording we

can distinguish 256 grey values for generating an image on a monitor. High-end digital cameras record in 16 bits, while older computer monitors can only support 8 bits per pixel.

Range (a) Distance. Radar RS distinguishes 'near range', 'far range', 'slant range', 'swath range', and 'ground range'.
(b) Interval.

Raster A set of regularly spaced (and contiguous) 2D 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.

Ray (a) A beam of radiant energy (*eg*, light). (b) Geometrically: any of a group of lines diverging from a common centre.

Reference plane In a topocentric coordinate system, the tangential plane to the Earth ellipsoid at the nadir of the image (thus defining a **3D rectangular terrain coordinate system**, the X-axis usually oriented eastward, the Y-axis northward, and the Z-axis upward perpendicular to the reference plane).

Reflectance The portion of the incident energy on a surface that is reflected; it is usually expressed as percentage. We call **spectral reflectance** the reflectance as a function of wavelength. Reflectance is sometimes also expressed as ratio with a value range 0 to 1 and then occasionally called reflectivity.

Reflection happens when a light ray meets a reflecting surface, such as a mirror. When light emitted by, *eg*, the Sun is directed onto the surface of

a mirror at a certain angle with respect to the surface normal (called incidence angle), it will be redirected into space at an angle, which is equal to the incidence angle (called the reflected angle).

Reflectivity is not a measure of the ability of reflective learning but a measure used in telecommunication and radar to quantify a target's reflection efficiency.

Refraction Light travels in a straight line unless it hits another medium. When light passes from one medium to another (*eg*, from air to water, or air to glass), it makes a deviation from its original straight path. This phenomenon is called refraction. Since the amount of deviation depends on the wavelength, we can see the effect of refraction in the form of a rainbow, or the splitting of white light into coloured light of the rainbow-spectrum as it passes through a glass prism.

Relief displacement Elevation dependent shift of an imaged object by a camera. The magnitude and direction of the shift does not only depend on elevation but also on the position of the object in the image, and the camera platform characteristics.

Remote sensing (RS) The art, science, and technology of observing an object, scene, or phenomenon by instrument-based techniques. 'Remote' because observation is done at a distance without physical contact with the object of interest.

Replicability (of image interpretation) refers to the degree of correspondence of image interpretation results obtained by different persons for the

same area or by the same person for the same area at different instants.

Resampling The process to generate a raster with another orientation or a different cell size than the original digital image and to assign DN-values using one of the following methods: nearest neighbour selection, bilinear interpolation, or bicubic interpolation.

Residual The difference between any measured quantity and the adjusted /computed value for that quantity.

Rods Photoreceptor cells in the retina of the eye for sensing brightness; see also **cones**.

Roughness A term with different meanings, among them the variation of surface elevation within a ground resolution cell. A surface appears rough to microwave illumination when the elevation variations become larger than a fraction of the radar wavelength.

Sampling (a) Selecting a representative part of a population for statistical analysis; to this end various strategies can be applied, such as random sampling, systematic sampling, stratified sampling, etc. In signal processing, (b) 'sampling' is the process of turning a continuous signal into a discrete one to achieve A/D conversion.

Satellite A manufactured object or vehicle intended to orbit the Earth, the moon, or another celestial body.

Scale A word of many different meanings. The map scale or photo scale is the ratio of a distance in the image and the corresponding horizon-

tal distance in the terrain. The ratio is commonly expressed as 1:m, where m is the **scale factor**. *Eg*, 1:25,000.

Scanner (a) 'Optical scanner', not a radio receiver scanning frequencies, nor a medial scanner, nor a desktop or photogrammetric scanner): an electro-optical remote sensor with only one or a few detectors and a scanning device. The most widely used scanning device is a moving mirror. Most scanners are multispectral scanners. A laser scanner is also an optical scanner, but an active sensor emitting and sensing monochromatic EM energy. (b) An office scanner or a photogrammetric scanner converts a hardcopy document, map, or photo to a digital image; these are typically 'flatbed scanners' using an entire linear CCD array as scanning device, which moves over the document that is mounted on a glass plate.

Scene Section of space and time in the real world.

Segmentation The process of dividing into segments. In **image segmentation** we aim at delineating regions that are homogeneous with respect to chosen spatial or radiometric characteristics. We can, *eg*, obtain a binary image (where each region consists exclusively of pixels with either the value 0 or the value 1) by thresholding the DNs, or we want to identify image segments such that each one is homogeneous according to a criterion on texture, etc.

Sensor In the context of this book, an instrument that detects and records EM energy. An **active sensor** is a device that generates itself the radiation it senses. A **passive sensor** detects radiation of an external source (solar, or terrestrial, or atmospheric radiation).

Slant range Distance as measured by the radar to each reflecting point in the scene and recorded in the side-looking radar image.

Spatial data In the broad sense, spatial data is any data with which position is associated.

Spatial resolution The degree to which an image can differentiate spatial variation of terrain features. Sometimes it is specified in the image space as pixel size, or lines per millimetre (lp/mm) for photographs. More relevant for applications is the specification in object space as ground sampling distance, or ground resolution cell size as determined by the IFOV.

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 interval of the EM spectrum to which the detector of a sensor is sensitive. The detector averages the spectral radiances within this range. A 'broadband sensor' such as the panchromatic camera of WorldView-1 averages per pixel the spectral response in the wavelength range from 0.4 to 0.9 μm. The spectral band of some SWIR cameras is 0.8 to 2.5 μm, while hyperspectral sensors have many but very narrow spectral bands.

Spectral reflectance curve The curve showing the portion of the incident energy that is reflected by a material as a function of wavelength.

Spectral resolution The degree to which the spectral response of a sensor is differentiated; specified as spectral band width.

Spectral response curve The curve portraying the sensitivity of a detector (*eg*, a CCD) to radiation per wavelength. The spectral sensitivity is a detector property and should not be confused with spectral reflectance curve, which portrays a terrain property.

Spectrometer A sensor, which measures radiance typically in many, narrow, contiguous spectral bands, usually in the visible to SWIR range of the spectrum; it offers a high spectral resolution, but low radiometric resolution as compared to a radiometer. Imaging spectrometers produce "hyperspectral images".

Specular reflection Mirror-like reflection; "bounced-off radiation".

Steradian (symbol sr) The SI unit of solid angle. It is used to describe two-dimensional angular spans in three-dimensional space, analogously to radian in a plane.

Stereo Short for stereoscopic. Stereoscopic viewing gives a three-dimensional impression. **Stereoscopy** is the science of producing three-dimensional visual models using two-dimensional images. We can make use of stereoscopy to make 3D measurements of objects.

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.

Stereograph A stereo pair arranged such (on the computer monitor, or on the table, or in a device) that you can readily get a 3D visual impression. Also referred to as 'stereogram'.

Stereoplotting The 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 taken from different positions. Stereoscopy is used in visual image interpretation and stereoscopic measurements (stereoplotting).

Stratified sampling Taking an equal amount of samples per strata; see also **sampling**.

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 Specification of a satellite orbit that is designed in such a way that the satellite always passes the same location on the Earth at the same local time.

Synthetic aperture radar (SAR) 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.

Terrain Terrain relief + terrain features (encompassing land and water).

Terrain coordinate system A system of expressing the position of a point in object space. Popular in EO are 3D rectangular coordinate systems (*eg*, a topocentric system; see also **reference plane**) and hybrid horizontal-vertical systems, where horizontal position is either defined by Lat-Lon or by 2D rectangular map coordinates.

Terrain elevation Elevation of a terrain point. A **terrain point** can be a point on the ground, or on a tree, a building, a water surface. Examples of terrain elevation data are the data sets stemming from SRTM or SPOT-5 HRS.

Terrain features Land cover, all kind of topographic objects that coincide with the ground surface or 'stick out' (the roads, buildings, trees, water bodies, etc), and any other characteristics of terrain except terrain relief.

Terrain relief Ground surface: its shape, not its composition nor its cover.

Terrain surface Envelop surface of terrain relief and terrain features (as represented by a DSM).

Texture A visual surface property; the word stems from weaving. Texture as an interpretation element expresses the spatial arrangement of tonal differences in an image.

Tone Among others one of the interpretation elements: the relative brightness in a black-and-white image.

Topography The description of a place where we live and move around.

Transmission The process of passing on energy through a medium or material.

Transmittance The amount of transmitted energy by the material under consideration, expressed either as percentage of transmitted to incident energy, or sometimes as ratio with a value range 0 to 1 and then occasionally called **transmissivity**.

Variable, interval A variable that is measured on a continuous scale. An interval variable is similar to an ordinal variable, except that the intervals between the values of the interval variable are equally spaced, so the difference between two values is meaningful. Different from a ratio variable it has no clear definition of zero, so it cannot be used to form ratios. *Eg*, temperature measured in Celsius (0 °C is not 'no temperature').

Variable, nominal A variable that is organized in classes, with no natural order, *ie*, cannot be ranked. It is also called a categorical variable.

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 clear definition of zero, so can be used to form ratios. *Eg*, temperature measured in Kelvin (0 K is 'no temperature').

Volume scattering The process of multiple reflections and redirection of radiation caused by heterogeneous material (atmosphere, water, vegetation cover, etc).

Wavelength The distance between two successive maxima of a periodic wave (in space); mostly stated in μm or nm.

Index

Appendix A

Spaceborne EO systems of special significance

A.1 Meteosat-8

The first Meteosat satellite was placed in orbit in 1977. Meteosat satellites are owned by the European organisation Eumetsat. Meteosat-8, launched in August 2002, also called MSG-1, is the first of the Meteosat Second Generation (MSG) satellites and, compared to the first generation, offers considerable improvements in spatial, spectral, radiometric and temporal resolution. In 2008, the pair Meteosat-8 and Meteosat-9 are in operation.

The spectral bands of the SEVIRI sensor (Table A.1) are chosen for observing phenomena that are relevant to meteorologists: a panchromatic band (PAN), mid-infrared bands, which give information about the water vapour present in the atmosphere, and thermal bands (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 and are used to detect thermal anomalies, such as forest fires or volcanic activity.

System	Meteosat-8
Orbit	Geo-stationary, 0° longitude
Sensor	SEVIRI (Spinning Enhanced VIS and IR Imager)
Swath width	Full Earth disc (FOV = 18°)
Off-nadir viewing	Not applicable
Revisit time	15 minutes
Spectral bands (μm)	0.5–0.9 (PAN), 0.6, 0.8 (VIS), 1.6 (NIR), 3.9, 6.2, 7.3, 8.7, 9.7 10.8, 12.0, 13.4 (TIR)
Spatial resolution	1 km (PAN), 3 km (all other bands)
Data archive at	www.eumetsat.de

Table A.1: Meteosat-8 SEVIRI characteristics.

System	NOAA-17
Orbit	812 km, 98.7° inclination, sun-synchronous
Sensor	AVHRR-3 (Advanced Very High Resolution Radiometer)
Swath width	2800 km (FOV = 110°)
Off-nadir viewing	No
Revisit time	2–14 times per day, depending on latitude
Spectral bands (μm)	0.58–0.68 (1), 0.73–0.98 (2), 1.58–1.64 (3A day), 3.55–3.93 (3B night), 10.3–11.3 (4), 11.5–12.5 (5)
Spatial resolution	1 km × 1 km (at nadir), 6 km × 2 km (at limb), IFOV=1.4 mrad
Data archive at	www.saa.noaa.gov

Table A.2: NOAA-17 AVHRR characteristics.

A.2 NOAA-17

NOAA stands for National Oceanic and Atmospheric Administration, which is a US-government body. The sensor onboard NOAA missions that is relevant for earth observation is the Advanced Very High Resolution Radiometer (AVHRR). Today, two NOAA satellites (NOAA-17, 18) are operational, older ones are standing by as a backup.

System	Landsat-7
Orbit	705 km, 98.2° inclination, sun-synchronous, 10:00 AM crossing, 16-day repeat cycle
Sensor	ETM+ (Enhanced Thematic Mapper)
Swath width	185 km (FOV = 15°)
Off-nadir 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 archives at	earthexplorer.usgv.gov edcimswww.cr.usgs.gov/imswelcome

Table A.3: Landsat-7 ETM+ characteristics.

A.3 Landsat-7

The Landsat programme is the oldest civil earth observation programme. It started in 1972 with the Landsat-1 satellite carrying the MSS multispectral sensor. In 1982, the Thematic Mapper (TM) replaced the MSS sensor. Both MSS and TM are scanners. In April 1999 Landsat-7 was launched carrying the ETM+ scanner. Today, only Landsat-5 and Landsat-7 are operational (Table A.3). Table A.4 lists example applications for the various bands.

Band	Wavelength (μm)	Example Applications
1	0.45–0.52 (Blue)	Coastal water mapping: bathymetry & quality Ocean phytoplankton & sediment mapping Atmosphere: pollution & haze detection
2	0.52–0.60 (Green)	Chlorophyll reflectance peak Vegetation species mapping Vegetation stress
3	0.63–0.69 (Red)	Chlorophyll absorption Plant species differentiation Biomass content
4	0.76–0.90 (NIR)	Vegetation species & stress Biomass content Soil moisture
5	1.55–1.75 (SWIR)	Vegetation-soil delineation Urban area mapping Snow-cloud differentiation
6	10.4–12.5 (TIR)	Vegetation stress analysis Soil moisture & evapotranspiration mapping Surface temperature mapping
7	2.08–2.35 (SWIR)	Geology: mineral and rock type mapping Water-body delineation Vegetation moisture content mapping
8	0.50–0.90 (15-m PAN)	Medium-scale topographic mapping Image sharpening Snow-cover classification

Table A.4: Example applications of the Landsat-7 ETM+ bands (after [15]).

A.4 Terra

EOS (Earth Observing System) is the centrepiece of NASA's Earth Science mission. The EOS AM-1 satellite, later renamed Terra , is the flagship of the fleet and was launched in December 1999. It carries five remote sensing instruments, including the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Advanced Spaceborne Thermal Emission and Reflectance Radiometer (ASTER). The ASTER instrument (Table A.5) is designed with three bands in the visible and near-infrared spectral range with a 15 m resolution, six bands in the shortwave infrared with a 30 m resolution, and five bands in the thermal infrared with a 90 m spatial resolution.

In addition to the Terra mission, the NASA has developed two other major satellites, called Aqua and Aura. While Terra is mainly focused on land applications, Aqua focuses on water, and Aura on trace gases in the atmosphere (specifically ozone). Aqua was launched in May 2002. Aura was successfully put in orbit in July 2004. Together, the satellites form a strong combination for earth observation, complementing each other with their data. The satellites are the backbone of what NASA calls its Earth Science Enterprise (ESE).

System	Terra
Orbit	705 km, 98.2° inclination, sun-synchronous, 10:30 AM crossing, 16-day repeat cycle
Sensor	ASTER
Swath width	60 km
Off-nadir viewing	across-track ± 8.5° SWIR and TIR ± 24° VNIR along-track 27.7° backwards band 3B
Revisit time	5 days (VNIR)
Spectral bands (μm)	VIS (bands 1–2) 0.56, 0.66, NIR 0.81 (3N nadir and 3B backward 27.7°), SWIR (4–9) 1.65, 2.17, 2.21, 2.26, 2.33, 2.40 TIR (bands 10–14) 8.3, 8.65, 9.10, 10.6, 11.3
Spatial resolution	15 m (VNIR), 30 m (SWIR), 90 m (TIR)
Data archives at	terra.nasa.gov edcimswww.cr.usgs.gov/imswelcome

Table A.5: Terra ASTER characteristics.

System	SPOT-5
Orbit	822 km, 98.7° inclination, sun-synchronous, 10:30 AM crossing, 26-day repeat cycle
Sensor	2× HRG (High Resolution Geometric) and HRS (High Resolution Stereoscopic)
Swath width	60 km
Off-nadir viewing	± 31° across-track
Revisit time	2–3 days (depending on latitude)
Spectral bands (μm)	0.50–0.59 (Green), 0.61–0.68 (Red), 0.78–0.89 (NIR), 1.58–1.75 (SWIR), 0.48–0.70 (PAN)
Spatial resolution	10 m, 5 m (PAN)
Data archives at	sirius.spotimage.fr www.vgt.vito.be (free VEGETATION data, older than 3 months)

Table A.6: SPOT-5 HRG characteristics.

A.5 SPOT-5

SPOT stands for Système Pour l'Observation de la Terre. The SPOT satellites are owned by a consortium of the French, Swedish and Belgium governments. SPOT-1 was launched in 1986. It was the first operational CCD line camera with across-track viewing capability to be put into space. At that time, the 10 m panchromatic spatial resolution was unprecedented in civil remote sensing. In March 1998 a significantly improved SPOT-4 was launched. Its HRVIR sensor has 4 instead of 3 bands, and the VEGETATION instrument was added. VEGETATION was designed for frequent (1–2 days revisit time) and accurate monitoring of the globe's landmasses at 1 km GSD. Some VEGETATION data is freely available from VITO, Belgium. SPOT-5 was launched in May 2002; it further offers improved spatial resolutions (Table A.6).

A.6 Satellites of IRS

The Indian Space Research Organisation (ISRO) launched in 1995 and 1997 two identical satellites, IRS-1C and IRS-1D within the Indian Remote Sensing (IRS) programme. IRS-1C carried three sensors, the Wide Field Sensor (WiFS), 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 with a high spatial resolution of 5.8 m. For a number of years, up to the launch of Ikonos in September 1999, the IRS-1C and -1D were the civil satellites with the highest spatial resolution.

In 2003, Resourcesat-1 was launched to continue the IRS-1C and 1D program.The specifications of LISS4 are listed in Table A.7. AWiFS has a GSD of 70 m. Resourcesat-1 has 60 Gigabits of onboard memory that allows for out-of-contact imaging.

Resourcesat-1 (IRS-P6), which is meant for agricultural applications, is one of several satellites for different applications:

- Oceansat-1 (IRS-P4) was launched in May 1999 to study physical and biological aspects of oceanography.
- The Technology Experiment Satellite (TES) was launched in October 2001. The satellite was intended to demonstrate and validate technologies that could be used in the future cartographic satellite missions of ISRO. TES carries a panchromatic camera with a spatial resolution of 1 m.
- Cartosat-1 (IRS-P5) was launched in 2005 with the purpose to produce stereo-images for mapping. It carries two panchromatic line cameras, one

System	Resourcesat-1
Orbit	817 km, 98.8° inclination, sun-synchronous, 10:30 AM crossing, 24-day repeat cycle
Sensor	LISS4
Swath width	70 km
Off-nadir viewing	± 20° across-track
Revisit time	5–24 days
Spectral bands (μm)	0.56, 0.65, 0.80
Spatial resolution	6 m
Data archive at	www.spaceimaging.com

Table A.7: Resourcesat-1 LISS4 characteristics.

forward looking and one backward looking. The nominal spatial resolution is 2.5 m, the effective resolution is around 3m. The swath width is 30 km.

- Cartosat-2 was launched in 2007. It has a single panchromatic camera capable of providing spot images for cartographic applications. The panchromatic camera is designed to provide images with better than one-meter spatial resolution and a swath of 10 km.

A.7 Ikonos

Ikonos was the first commercial "high resolution satellite" to be placed into orbit. Ikonos is owned by GeoEye (merger of SpaceImaging and OrbImage), a USA-based earth observation company). Other commercial high resolution satellites are GeoEye-1 (OrbView-5, launched in 2008; the GSD of the pan mode is 0.4 m), QuickBird (launched in 2001, owned by DigitalGlobe; the GSD of the pan mode is 0.6 m), and EROS-A1 (launched in 2000). Ikonos was launched in September 1999 and regular data ordering has been taking place since March 2000.

The OSA sensor onboard Ikonos (Table A.8) is based on the pushbroom principle and can simultaneously take panchromatic and multispectral images. In addition to the high spatial resolution of 1 m panchromatic and 4 m multispectral, it also has a high radiometric resolution using 11 bits quantization.

System	Ikonos
Orbit	681 km, 98.2° inclination, sun-synchronous, 10:30 AM crossing, 14-day repeat cycle
Sensor	Optical Sensor Assembly (OSA)
Swath width	11 km
Off-nadir viewing	± 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 A.8: Ikonos OSA characteristics.

System	EO-1
Orbit	705 km, 98.7° inclination, sun-synchronous, 10:30 AM crossing, 16-day repeat cycle
Sensor	Hyperion
Swath width	7.5 km
Off-nadir viewing	?
Revisit time	16 days
Spectral bands	220 bands, covering 0.4 μm to 2.5 μm
Spatial resolution	30 m
Data archive at	eo1.gsfc.nasa.gov

Table A.9: EO-1 Hyperion characteristics.

A.8 EO-1

The Earth Observing-1 (EO-1) mission was part of the NASA New Millennium Program, which is focused on new sensor and spacecraft technologies that can directly reduce the cost of Landsat and related earth monitoring systems. EO-1 was launched in 2000 and is still in operation.

The EO-1 satellite was originally in an orbit that covered the same ground track as Landsat 7, approximately one minute later. This enabled EO-1 to obtain images of the same ground area at nearly the same time, so that direct comparison of results could be obtained from Landsat ETM+ and the three primary EO-1 instruments. The three primary instruments on the EO-1 spacecraft are the Hyperion, the LEISA Atmospheric Corrector (LAC), and the Advanced Land Imager (ALI). LEISA is the Linear Etalon Imaging Spectrometer Array.

Hyperion is a 220-band 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 nm to 2500 nm. The specifications of EO-1/Hyperion are summarized in Table A.9.

LAC is an imaging spectrometer covering the spectral range from 900 nm to 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 EO-1 Advanced Land Imager (ALI) is a technology verification instrument. Operating in a pushbroom fashion at an orbit of 705 km, the ALI provides Landsat-

System	Proba micro-satellite (94 kg)
Orbit	615 km, 97.9° inclination, sun-synchronous, 10:30 AM crossing, 7-day repeat cycle
Sensor	CHRIS (Compact High-Res. Imaging Spectrometer)
Swath width	14 km
Off-nadir viewing	along-track $\pm$ 55°, across-track $\pm$ 36°
Revisit time	less than 1 week, typically 2–3 days
Spectral bands	19 or 63 bands, 410 nm to 1050 nm
Spatial resolution	18 m (full spatial resolution), 36 m (full spectral resolution)
Data archive at	www.chris-proba.org.uk

Table A.10: Proba CHRIS characteristics.

type panchromatic and multispectral bands. These bands have been designed to mimic six Landsat bands with three additional bands covering 0.433 μm to 0.453 μm, 0.845 μm to 0.890 μm, and 1.20 μm to 1.30 μm.

A.9 Proba/CHRIS

ESA's micro-satellite Proba (Project for On-Board Autonomy), launched in October 2001, is a technology experiment to demonstrate the onboard autonomy of a generic platform suitable for small scientific or application missions. Proba carries several instruments. One of these is CHRIS, the Compact High Resolution Imaging Spectrometer (Table A.10). CHRIS is used to measure directional spectral reflectance of land areas, thus providing new biophysical and biochemical data.

It appears that the reflectance of most materials depends on the angle of incidence, the angle of reflectance, and the two azimuthal angles. This dependency is captured by what is called the Bidirectional Reflectance Distribution Function (BRDF), which fully describes the directional dependency of the reflected energy. Measuring the BRDF may give us clues about the characteristics of the observed materials or objects.

In addition to CHRIS, Proba carries a panchromatic High-Resolution Camera (HRC) with a spatial resolution of 5 m.

Envisat-1	ASAR	GOMOS	RA-2	MERIS	MIPAS	MWR	LR	SCIAMACHY	DORIS	AATSR
Atmosphere										
Clouds				•				•		•
Humidity						•				
Radiative Fluxes				•						•
Temperature					•			•		
Trace Gases		•			•			•		
Aerosols		•		•	•			•		•
Land										
Surface Temperature										•
Vegetation Characteristics	•			•						•
Surface Elevation	•		•				•		•	
Ocean										
Ocean Colour				•						
Sea Surface Temperature										•
Surface Topography			•				•			
Turbidity				•						
Wave Characteristics	•		•							
Marine Geoid			•							
Ice										
Extent	•			•						•
Snow Cover	•			•						•
Topography	•		•				•			
Temperature										•

Table A.11: Instruments of Envisat-1 listed with their applications.

A.10 Envisat-1

Envisat-1 was launched in 2001 to provide measurements of the atmosphere, ocean, land and ice. Envisat-1 is a large satellite (8200 kg, in-orbit configuration 26 m $\times$ 10 m $\times$ 5 m) carrying an array of different sensors. It was designed to provide synergy between various scientific disciplines, thus making the total payload complement more than just the sum of the instruments (Table A.11).

The Envisat-1 satellite comprises a set of seven ESA developed instruments supported by three complementary instruments (AATSR, SCIAMACHY and DORIS):

- ASAR, Advanced Synthetic Aperture Radar.
- MERIS, Medium Resolution Imaging Spectrometer.
- RA-2, Radar Altimeter 2.
- MWR, Microwave Radiometer.
- LR, Laser Retro-Reflector.
- GOMOS, Global Ozone Monitoring by Occultation of Stars.
- MIPAS, Michelson Interferometer for Passive Atmospheric Sounding.
- AATSR, Advanced Along Track Scanning Radiometer.
- DORIS, Doppler Orbitography and Radiopositioning Integrated by Satellite.
- SCIAMACHY,
  Scanning Imaging Absorption Spectrometer for Atmospheric Cartography.

System	Envisat-1
Orbit	800 km, 98.6° inclination, sun-synchronous, 10:00 AM crossing, 35-day repeat cycle
Sensor	ASAR (Advanced SAR)
Swath width	56 km to 405 km
Off-nadir viewing	across-track 17° to 45°
Revisit time	35 days
Frequency	C-band, 5.331 GHz
Polarization	several modes: HH+VV, HH+HV or VV+VH
Spatial resolution	30 m or 150 m (depending on mode)
Sensor	MERIS
Swath width	1150 km
Revisit time	3 days
Spectral range (μm)	0.39–1.04 (VNIR)
Spectral bandwidth	1.25 nm to 25 nm (programmable)
Bands	15 bands (due to limited capacity)
Spatial resolution	300 m (land), 1200 m (ocean)
Data archive at	envisat.esa.int

Table A.12: Characteristics of the Envisat-1 satellite and its ASAR and MERIS sensors.

Additionally, ESA's Artemis data relay satellite system is used for communication to the ground.

The most important sensors for land applications are the Advanced Synthetic Aperture Radar (ASAR), the Medium-Resolution Imaging Spectrometer (MERIS), and the Advanced Along-Track Scanning Radiometer (AATSR). The characteristics of the Envisat-1 satellite and its ASAR and MERIS sensors are listed in Table A.12.

Envisat-1's ASAR ensures the continuation of the ERS-1 and ERS-2 radar satellites. It features enhanced capabilities in terms of coverage, range of incidence angles, polarizations and modes of operation. In normal image mode, the ASAR generates high-spatial resolution products (30 m) similar to the ERS SAR products. It can image seven different swaths located over a range of incidence angles from 15° to 45° in HH or VV polarization. In other modes the ASAR is capable of recording the cross-polarization modes HV and VH. In addition to the 30 m resolution, it offers a wide swath mode, for providing images of a wider strip (405 km) with a medium resolution (150 m).

The Medium Resolution Imaging Spectrometer Instrument, MERIS, is a 68.5° FOV pushbroom imaging spectrometer that measures the solar radiation reflected by the Earth, at a ground spatial resolution of 300 m, in 15 spectral bands in the visible and near infra-red. In fact, the bands of MERIS are fully programmable in width and position, making MERIS a fully-fledged imaging spectrometer. However, due to the capacity constraints on the entire Envisat-1

Band	Band centre (nm)	Band width (nm)	1c|Potential Applications
1	412.5	10	Yellow substance, turbidity
2	442.5	10	Chlorophyll absorption maximum
3	490	10	Chlorophyll and other pigments
4	510	10	Turbidity, suspended sediment, red tides
5	560	10	Chlorophyll reference, suspended sediment
6	620	10	Suspended sediment
7	665	10	Chlorophyll absorption
8	681.25	7.5	Chlorophyll fluorescence
9	705	10	Atmospheric correction, vegetation
10	753.75	7.5	Oxygen absorption reference, vegetation
11	760	2.5	Oxygen absorption band
12	775	15	Aerosols, vegetation
13	865	20	Aerosols corrections over ocean
14	890	10	Water vapour absorption reference
15	900	10	Water vapour absorption, vegetation

Table A.13: Envisat-1 MERIS band characteristics.

system, a choice was made to offer a "standard" set of 15 bands (Table A.13). MERIS allows global coverage of the Earth in 3 days. The MERIS instrument can be operated either in direct or in averaging observation modes. In averaging mode, the data are spatially averaged onboard to produce a 1200 m resolution image. In direct mode the instrument delivers the full resolution product of 300 m resolution. Typically, the direct mode is used on land and coastal features, while the averaging mode is used over the ocean. The full spatial resolution is available for up to 20 minutes per orbit, or about 20% of the time.

A.11 Future development

The following list is compiled based on current initiatives.

- *Fast development*. Nowadays, the time from the drawing board to the launch in space of a remote sensing satellite can be as short as one year. This means that satellites can be developed and deployed very fast.
- *New sensor types*. We may see a development of sensor types that were previously unused in space. Examples are hyperspectral sensors and P-band radar (see the chapter on active sensors).
- *New attitude control methods*.Traditionally, a large part of the mass of a satellite consists of fuel needed for attitude control and orbital corrections. The use of new electric propulsion methods, such as magnetic torque and ion motors, may drastically reduce the mass of a satellite, thus opening the road to miniaturization.
- *Better pointing capabilities*. Many of the current satellite have improved pointing capabilities, which allows faster revisits. This means that the same area can be observed twice within a couple of days rather than a couple of weeks.
- *Big versus small*. Nowadays, satellites come in all sizes ranging from the huge Envisat-1, with a mass larger than 8000 kg, to the tiny SNAP-1, with a mass as little as 6 kg. Currently there seem to be two lines of developments. One is the development of large, multi-instrument satellites, such

as Terra and Envisat-1. The other is the development small, mostly single-instrument satellites. Mini satellites can be developed and launched at a low cost. For the price of one large satellite, one can have a *constellation* of small satellites (see below).

- *High-speed laser communication*. One of the bottlenecks in remote sensing has always been the capacity of the downlink channel. Basically, the amount of data that could be sent to the ground was limited by the microwave channels that were used. The use of modern laser optical communication channels may give rise to an unprecedented flow of data to the Earth. Already SPOT-4 and Envisat-1 were equipped with experimental laser-optical communication instruments to send data via ESA's geostationary communication satellite Artemis.

- *Special missions*. Instead of having one multi-purpose satellite, there may be many specialized remote sensing satellites. For instance, the British-Chinese satellite Tsinghua-1, launched in 2000, was developed to provide daily world-wide high-resolution imaging for disaster monitoring and mitigation. Tsinghua-1 is the first demonstrator for the Disaster Monitoring Constellation (see next item).

- *Constellations of satellites*. AlSAT-1, the 90 kg Algerian microsatellite, which was launched in November 2002, is part of the first dedicated Disaster Monitoring Constellation (DMC). The DMC comprises seven earth observation microsatellites launched into low-earth orbit to provide daily imaging revisit anywhere in the world.

- *New spacefaring countries*. The DMC Consortium comprises a partnership between organizations in Algeria, China, Nigeria, Thailand, Turkey, Viet-

nam and the United Kingdom. Six of the seven microsatellites for the DMC are constructed at SSTL (Surrey Satellite Technology Limited) in the UK. The seventh microsatellite (Thai-Paht-2) is built at the Mahanakorn University of Technology (MUT) in Bangkok, Thailand.

- *Use of 'new' orbits*. In the future, satellites may be placed in orbits that were not used before for earth observation, such as the Lagrangian points of the Earth-Moon system, or the Sun-Earth system.

- *Improved coverage*, achieved through having more satellites, more receiving stations, higher onboard storage capacity, better data throughput, better pointing capabilities, etc.

- *Rapid dissemination*. By exploiting the capabilities of modern computing systems and networks, remote sensing data can be distributed to the end-user almost in real-time.

- *Cheap & useful data*. Stiff competition on the remote sensing market has caused prices to drop significantly. On the other hand, the quality of free data sources has improved considerably. When starting a project, it is advisable to check the free and low-cost data sources first before spending a huge amount of money on commercial data.

Appendix B

SI units & prefixes

Quantity	SI unit
Length	metre (m)
Time	second (s)
Temperature	kelvin (K)
Energy	joule (J)
Power	watt (W) (J/s)
Pressure	pascal (Pa)
Frequency	hertz (Hz)
Angle	radian (rad)
Solid angle	steradian (sr)

Table B.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 B.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 B.3: Common units of length.

Parameter	Value
speed of light	$2.9979 \cdot 10^{8}$ m/s
degree celcius ($^{\circ}$C)	(degree celcius + 273.15) K
inch	2.54 cm
foot	30.48 cm
mile	1.609 m
millibar	1 hPa

Table B.4: Constants and non-SI units.

Appendix C

List of Formulae

2.1 Sine wave
2.2 Speed of light in function of wavelength and frequency
2.3 Energy held by a photon
2.4 Radiance at the detector

4.1 Size of the GRC for a digital camera
4.2 Diameter of GRC of a scanner

5.1 Intensity in function of red, green, and blue
5.2 Gain of a filter
5.3 Brovey transform

5.4 Pixel addition in image fusion

6.1 Relief displacement
6.2 1^{st} order transformation for x
6.3 Affine transformation, together with (6.2)
6.4 RMSE in x-direction
6.5 Overall RMSE

9.1 Photo scale in function of flying height and focal length

10.1 Radar equation
10.2 Laser range

12.1 Planck's radiation law
12.2 Wien's displacement law
12.3 Stefan-Boltzmann law
12.4 Emissivity
12.5 Kirchhoff's law
12.6 Radiant temperature in function of kinetic temperature
12.7 Radiant temperature at the top of the atmosphere
12.8 Radiation observed by the remote sensor

13.1 Binary encoding for spectral matching
13.2 Vector angle between two spectra in vector notation