The core of GIScience

a process-based approach

ITC Educational textbook series

UNIVERSITY OF TWENTE
FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

ITC

Chapter 4

Sensors

Wim Bakker
Wan Bakx
Wietske Bijker
Karl Grabmaier
Lucas Janssen
John Horn
Gerrit Huurneman
Freek van der Meer
Christine Pohl
Klaus Tempfli
Valentyn Tolpekin
Tsehaie Woldai

4.1 Platforms and passive electro-optical sensors

Having explained the physics of sensing in Chapter 2, in this chapter we discuss sensor systems and set out to discover the logic of current electro-optical sensing technology. First, in Subsection 4.1.1, we will look at the characteristics of platforms used for geospatial data acquisition (GDA) from the air and from space: various platforms such as aircraft, space shuttles, space stations and satellites are used to carry one or more sensors for Earth Observation. Next, Subsection 4.1.2 will elaborate on frame and line cameras; the latter, which can be operated from the air or space, are also known as *pushbroom sensors*. Optical scanners (also referred to in the literature as across-track scanners or *whiskbroom scanners*) are treated in Section 4.1.3, which discusses multispectral, hyperspectral and thermal scanners in detail. Some camera systems can provide us with *stereo* images, justifying a short introduction to stereoscopy in Subsection 4.1.4.

4.1.1 Platforms and missions

Sensors used in Earth Observation can be operated at altitudes ranging from just a few centimetres above the ground—using field equipment—to those 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 could mount a spectrometer on a pole to measure the changing reflectance of a specific crop during the day or over a whole season.

Moving platforms

To gain a wider view, we use aircraft at altitudes ranging up to approximately 20 km. Depending on the type of aerial survey and the weight of equipment and survey costs, we can choose from a variety of vehicles. Fixed-wing aircraft are used for thermal scanning and a systematic photo-coverage for topographic mapping, land titling projects, and the like. Aerial survey cameras are heavy and they are fixed to a stabilized mount set in a hole in the floor of the aircraft. Most survey airplanes fly lower than 8 km but higher than 1000 m. They can fly as slow as 150 km h^{-1}, but even at that speed image quality is already affected by motion blur unless the camera is fitted with a compensation device. Aerial survey cameras are highly sophisticated and expensive.

Airborne laser-scanner systems used to be heavy, but nowadays the total weight of the equipment can be as light as 30 kg. Laser scanners are either mounted on fixed-wing aircraft or helicopters, the latter being able to fly very slowly at low altitudes, thus allowing the acquisition of highly detailed data (at high costs per unit of area). The small-format cameras used are cheaper and lighter than large-format aerial survey cameras, making it possible to mount these systems on micro-light airplanes for urban reconnaissance, or even kites (e.g. for surveying an industrial area). Unmanned aerial vehicles (UAVs) are gaining popularity for observing dangerous areas or to reduce costs. A special type of UVA, the High Altitude Long Endurance (HALE) vehicle, can bridge the gap between manned survey aircraft and spacecraft or satellites. Typically, a HALE is a remotely operated aircraft of ultra-light weight and load that flies for months at altitudes of around 20 km.

A key advantage of aerial surveys is that they can be "targeted". The survey can be undertaken at exactly the required time and can be done with exactly the required spatial resolution by having the aircraft fly at the required altitude. Moreover, in comparison with civilian satellite RS, we can acquire images of much higher spatial resolution, enabling recognition of objects of much smaller size. With current aerial survey cameras, we can achieve a pixel size on the ground as small as 5 cm.

Satellites are launched by rocket into space, where they then circle the Earth for 5 to 12 years on a predefined orbit. The choice of orbit depends on the objectives of the sensor mission; orbit characteristics and different orbit types are explained below. A satellite must travel at high speed to orbit at a certain distance from the Earth; the closer to the Earth, the faster the speed required. A space station such as ISS has a mean orbital altitude of 400 km and travels at roughly 27,000 km h^{-1}. The Moon at a distance of 384,400 km can conveniently circle the Earth at only 3700 km h^{-1}. At altitudes of 200 km, satellites already encounter traces of the atmosphere, which causes rapid orbital and mechanical decay. The higher the altitude, the longer is the expected lifetime of the satellite.The majority of civilian Earth-observing satellites orbit at altitudes ranging from 500 to 1000 km. Here we generally find the "big boys", such as Landsat-7 (2200 kg) and Envisat (8200 kg), but the mini-satellites of the Disaster Management Constellation (DMC) also orbit in this range. DMC satellites have a weight of around 100 kg and were launched by several countries into space early in the current millennium at relatively low-cost. These satellites represent a network for disaster monitoring that provides images in three or four spectral bands with a ground pixel size of 32 m or smaller.

satellites

Satellites have the advantage over aerial survey of continuity. Meteosat-9, for example, delivers a new image of the same area every 15 minutes and it has done so every day for many years. The high temporal resolution at low cost goes together with a low spatial resolution (pixel size on the ground of 1×1 km^2). Both the temporal and the spatial resolution of satellite remote sensors are fixed. While aerial surveys have been restricted in some countries, access to satellite RS data is commonly easier, although not every type of satellite RS image is universally available.

Aerial survey missions

direct sensor orientation

Modern airborne sensor systems use a high-end GPS receiver and many also include an Inertial Measuring Unit (IMU). GPS is used for navigation and for coarse "sensor positioning". We need to know the coordinates of the exposure stations of a camera to relate points and features in the images to positions on the ground; differential GPS is applied 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 IMU has two advantages: IMU readings can be used to improve the accuracy of the coordinates obtained by GPS (achieving a RMSE better than 0.1 m); and the IMU measures the attitude angles of the sensor (Figure 4.27). An IMU, an assemblage of gyros and accelerometers, is a sophisticated, heavy, and expensive instrument that was originally used only in Inertial Navigation Systems (INSs). Measuring continuously the position and attitude of the moving sensor, an IMU allows us to relate the sensor recordings to position in the terrain in near 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 also solve the georeferencing problem indirectly (see Section 5.3).

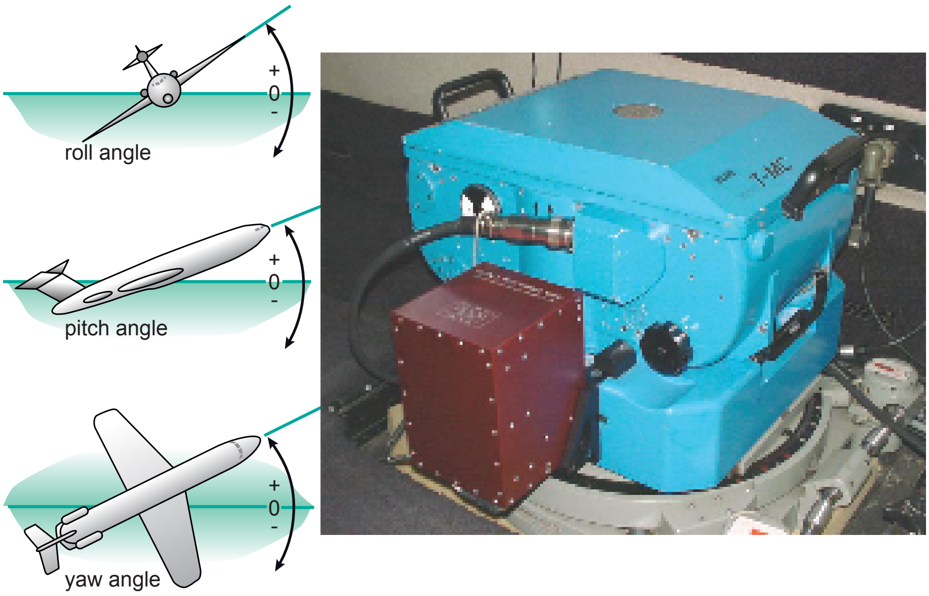

Figure 4.1
Attitude angles (left) and an IMU attached to a Zeiss RMK-TOP aerial camera (courtesy of IGI).

Mission planning and execution is usually done by commercial survey companies or, otherwise, by large national mapping agencies or the military. During missions, companies use professional software for flight planning and, most likely, one of the two integrated aircraft guidance and sensor management systems available (produced by APPLANIX or IGI). Pioneering work on computer-controlled navigation and camera management was done at ITC in the days when it still had an aerial photography and navigation department. The basics of planning aerial survey missions are explained in Section 4.6.

Satellite missions

orbit parameters

The monitoring capabilities of a satellite-borne 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 around 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, 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 (i.e. the *spatial coverage*) and the details that can be observed (i.e. the *spatial resolution*). In general, the higher the altitude, the larger 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 orbits the Earth between the latitudes 60 ° N and 60 ° S. 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 Earth's polar regions.

- *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 for 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 resulting from the high speed. Short exposure times, however, require high intensities of incident radiation, which is a problem in space because of atmospheric absorption. It should be obvious that the contradictory demands of high spatial resolution, no motion blur, high temporal resolution, long satellite lifetime (thus lower cost) represent a serious challenge for satellite-sensor designers.

- *Repeat cycle* is the time (in days) between two successive identical orbits. The *revisit time* (i.e. 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, and not only vertically downwards. Many 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 Subsection 4.1.4).

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

- *Polar orbit* refers to orbits with an inclination angle between 80 ° and 100 °. An orbit having an inclination larger than 90 ° means that the satellite's motion is in a westward direction. Such a polar orbit enables observation of the whole globe, also near the poles. Satellites typically orbit at altitudes of 600–1000 km.

- *Sun-synchronous orbit* refers to a polar or 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 mid-morning, at around 10:30 h local solar time. At that moment the Sun angle is low and the shadows that creates reveal terrain relief. In addition to day light images, a Sun-synchronous orbit also allows the satellite to record night images (thermal or radar, passive) during the ascending phase of the orbit on the night side of the Earth.

- A *Geostationary orbit* refers to orbits that position the satellite 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 has a fixed position relative to the Earth. Geostationary orbits are used for meteorological and telecommunication satellites.

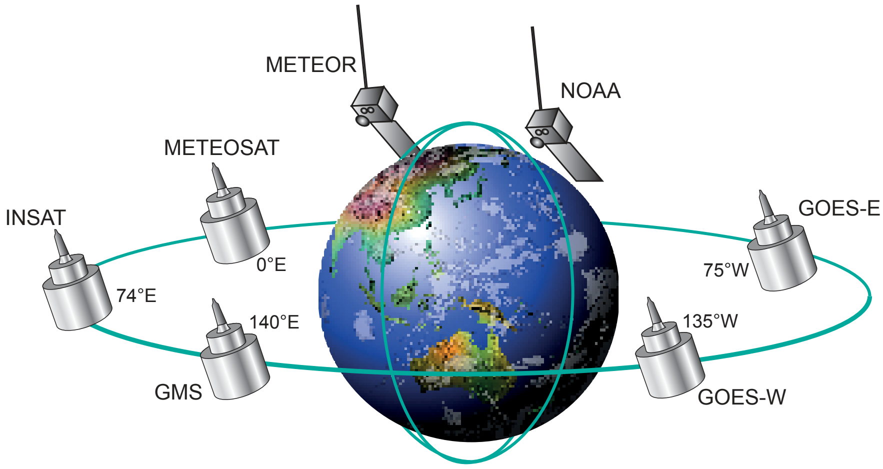

Figure 4.2
Meteorological observation by geostationary and polar satellites.

Today's meteorological weather satellite systems use a combination of geostationary satellites and polar orbiters (Figure 4.28). 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 parameters to facilitate georeferencing of the images. High resolution sensor systems such as Ikonos or QuickBird use GPS receivers and star trackers as their POS.

The data from space-borne sensors need to be transmitted to the ground in some way. Russia's SPIN-2 satellite, with its KVR camera, used film cartridges that were dropped over a designated area on the Earth. Today's 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 current trend is that small receiving units, consisting of a small dish with a PC, are being developed for local reception of RS data.

4.1.2 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 three channels, one for each primary colour (red, green, blue); three elongated CCDs next to each other constitute

CCD

one square "colour pixel". Each CCD has its colour filter right on top to only transmit the required band of incident light. 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, an aircraft or a spacecraft. SPOT-1, launched in 1986, was the first satellite to use a line camera. Line cameras build up a digital image of an area line by line (Figure 4.4). In the older literature, therefore, it is also referred to as *pushbroom scanner*, as opposed to a *whiskbroom scanner* (see Subsection 4.1.3), which actually scans (across the track of the moving platform).

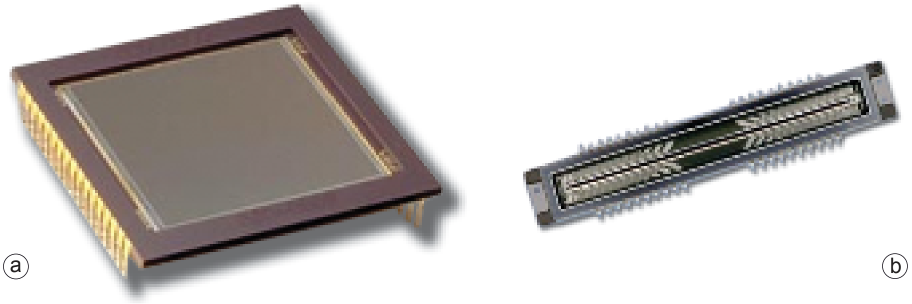

Figure 4.3
Two CCD chips: (a) matrix array Kodak KAF-16801, pixel size 9 µm; (b) linear array Kodak KLI-14403, pixel size 5 µm.

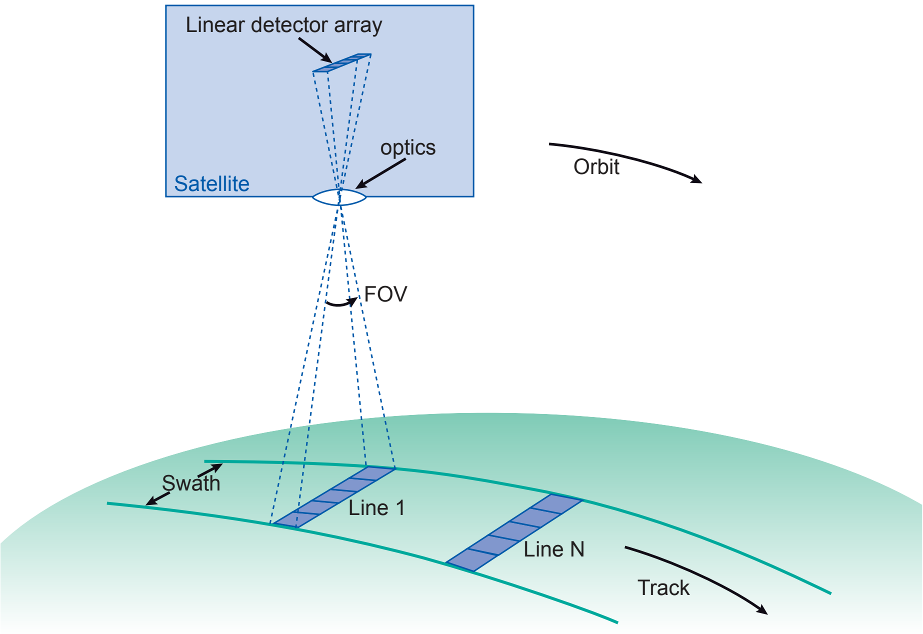

Figure 4.4
Principle of imaging by line camera on a spacecraft ('pushbrooming').

Detector arrays

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

The spectral sensitivity of a sensor band is commonly specified by a lower- and an upper-bound wavelength of the spectral band covered, e.g. 0.48 to 0.70 µm for the spectral sensitivity

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 chosen at the wavelengths where the 50% response is achieved. The DN produced by a detector results from averaging the spectral response of incident radiation. Figure 4.5 shows 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 show, moreover, varying degrees of degradation over 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 Section 5.2).

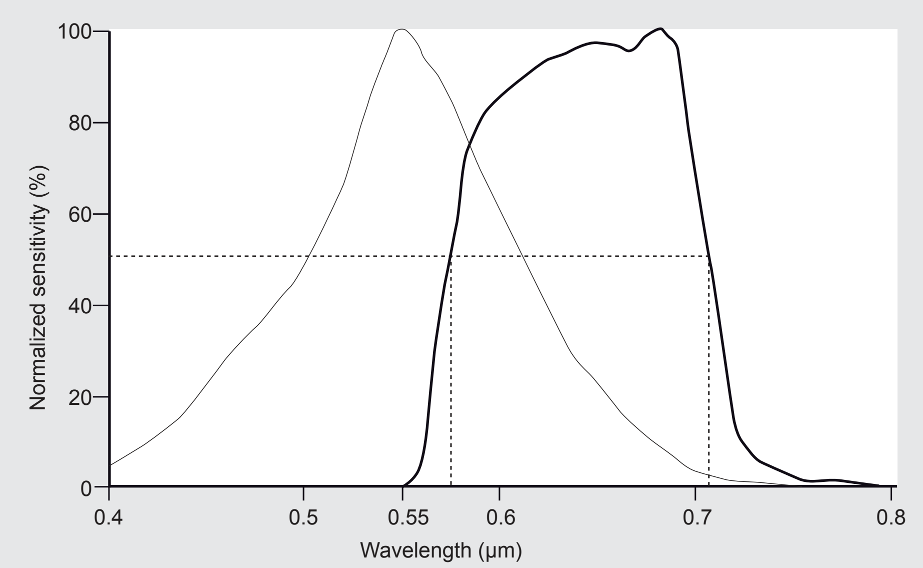

Figure 4.5
Normalized spectral response curve of (a) channel 1 of NOAA's AVHRR and (b) the spectral sensitivity of the rods of the human eye.

When compared with photographic film, most CCDs have a much higher general sensitivity and thus they need less light. The reason is that 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.

dynamic range

If we were interested in a high radiometric resolution, we would like a CCD to have a wide dynamic range. *Dynamic range* is the ratio of the maximum to the minimum level of intensity that can be measured; it is also known as the signal to noise ratio of the detector. The maximum intensity is determined by the maximum charge capacity of the semiconductor cell. The minimum intensity is determined by the noise level. Noise is unwanted collected charge, for example caused by unblocked IR or UV radiation for a CCD that should be sensitive to blue light. It only makes sense to record a DN of many bits if the semiconductor cell has a wide dynamic range. It is the manufacturer's concern to ensure sufficient dynamic range to meet the radiometric resolution (expressed in bits) required by the user. 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 pockets before we had any digital airborne camera offering satisfactory surveying quality. The main reason for this is the ultra-high 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 (e.g. for red, green, blue, NIR). ASTER has two panchromatic channels, one linear array looking vertically down (nadir view) and the second looking backwards; the two resulting images 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 consists of 12,000 CCDs.

linear arrays

Current CCD technology enables the production of very high quality linear arrays but not (yet) the very large matrix arrays that would be needed for large-format digital aerial cameras to be able to match the well-proven film-based survey camera. The two market leaders in digital aerial frame cameras, ZI and Microsoft (former Vexcel), therefore use several detector arrays for panchromatic imaging and software to compile a single large-format image from the sub-frames. ZI's DMC has, for example, 13,500 × 7,500 CCDs per sub-frame. One of the advantages of 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 on the market.

matrix arrays

Optical system

Cameras use either lenses or telescopes to focus incident radiation onto 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). The camera head of a digital aerial frame camera (such as Z/I's DMC and Vexcel's UltraCam) even consists of several of such lenses to focus the light rays on the respective CCD arrays. However complicated a lens may be physically, geometrically 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 its *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 the 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. At a flying height of 2000 m, we would attain a ground-resolution cell size in the across-track direction of 21 cm, provided the airplane flies perfectly horizontally over flat terrain (see Figure 4.6). You would conclude correctly that the ADS40 has a CCD/pixel size of 6.5 µm. The ratio $\frac{H}{f}$ is referred to as the *scale factor* of imaging.

Space-borne cameras do not have lens cones—they have telescopes. The telescopes of Ikonos and QuickBird consist of an assembly of concave and flat mirrors, thus achieving a spatial resolution that 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 nadir.

telescope

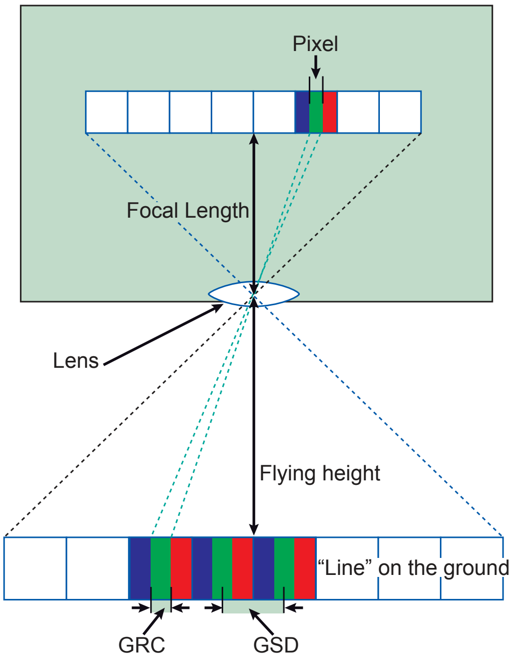

Figure 4.6
Pixel, ground resolution cell, ground sampling distance for digital cameras.

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 (e.g. in the case of a nadir view of horizontal terrain); see Figure 4.6. If a space-borne 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, Formula 4.1 does not apply anymore; the ground resolution cell size and the GSD increase with increasing distance from nadir. Section 5.3 explains how to deal with this.

ground resolution cell

ground sampling distance

Digital aerial cameras have several advantages over film cameras, pertaining to both the quality of images and economics. Digital aerial cameras commonly record in 5 spectral bands (panchromatic, RGB, NIR), therefore, we can obtain with one flight panchromatic stereo images, true colour images and false colour images; with a film camera we would have to fly this course three times and develop three different types of film. The radiometric quality of CCDs is better than that of photographic film. Digital cameras also allow an all-digital workflow, making processing faster and cheaper. Digital cameras can acquire images with a high likelihood of redundancy without additional costs for material and flying time; this favours automated information extraction. Finally, new geoinformation products can be generated as a result of various extended camera features. In Subsection 4.1.3 multispectral scanners are introduced. Line cameras as compared to across-track scanners have the advantage of better geometry. Airborne line cameras and scanners require gyroscopically stabilized mounts to reduce any effects of aircraft vibration and compensate for rapid movements of the aircraft. Such a stabilized mount keeps a camera in an accurate level position so that

advantages of digital cameras

it continuously points vertically downward. We want vertical images for mapping because of the better geometry. we, Therefore, also mount large-format digital frame cameras and film cameras on stabilized platforms for applications that require high-quality images.

4.1.3 Scanners

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 (e.g. 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, a rotating mirror, or a nutating mirror. An alternative, which is used for laser scanning, is fiber optics.

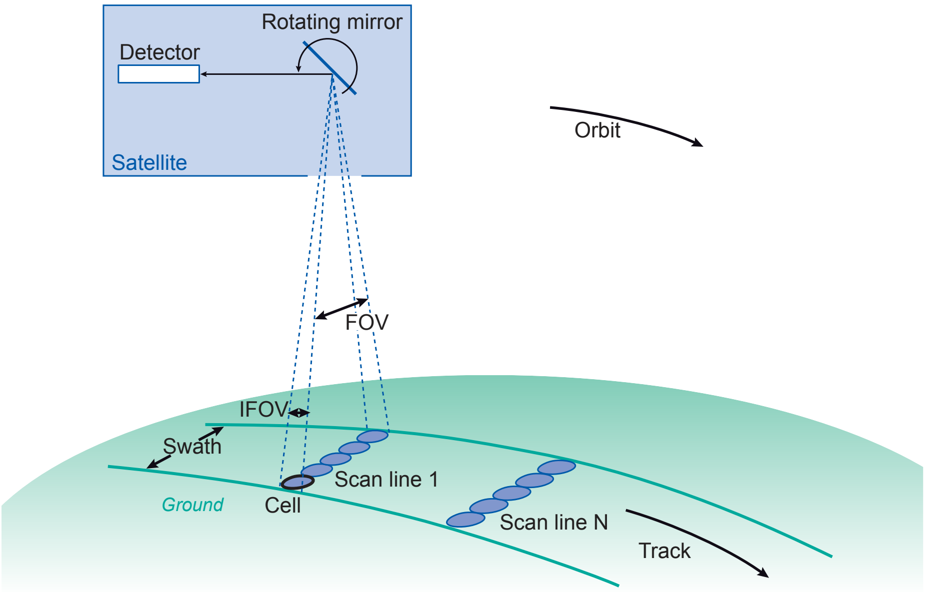

Figure 4.7
Principle of an across-track scanner.

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 (e.g. 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 scanner. A multispectral scanner has at least one detector per spectral band. Different from small-format frame cameras, for which filters are used to separate wavelength bands, scanners and line cameras use a prism and/or a grating as a *beam splitter*. A *grating* is a dispersion device used for splitting up SWIR and TIR radiation. Hyperspectral scanners also use gratings. A *prism* can split higher frequency radiation into red, green, blue, and NIR components. A simple

RGB and NIR scanner produces in one sweep of the mirror a single image line for each of the four channels.

Instead of using only one detector per band, space-borne scanners use several. The first civil space-borne remote sensor, Landsat MSS (launched in 1972), used six per band (thus, in total, 24; see Figure 2.18). ASTER uses 10 detectors for each of its five TIR channels. One sweep of the 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. Section 5.2 treats the correcting of an image for periodic *line dropouts*.

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 nadir, the cell is circular, of diameter D. D depends on the IFOV, β, of the detector and the flying height.

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

A scanner with β = 2.5 mrad operated at H = 4000 m would have, therefore, a ground resolution of 10 m at nadir. Towards the edge of a swath, the ground resolution cell becomes elongated and bigger (Figure 4.29).

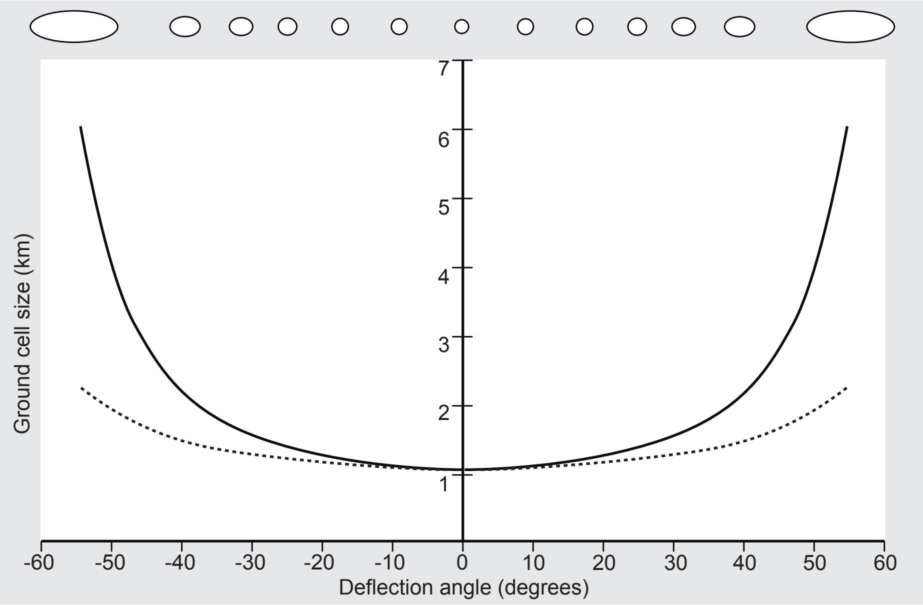

Figure 4.8
GRC of NOAA's AVHRR: at nadir the cell diameter is 1.1 km; at the edge the ellipse stretches to 6.1×2.3 km. The solid line shows the across-track resolution, the dashed line the along-track resolution. The ellipses at the top show the shape of the ground cells along a scanned line. NOAA processes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

The width of the area that 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 in the AVHRR design. Landast-7 has an 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 the 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 greater than or equal to D across the track).

4.1.4 Stereoscopy

Stereoscopy, the science of producing three-dimensional (3D) visual models using two-dimensional (2D) images, dates back to the 16th century. The astronomer Kepler was presumably the first person to define stereoscopic viewing. One of the main reasons for being able to perceive depth is that we have two eyes, which enables 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 being partially obscured by the other one, or one appears smaller than the other although they are 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 4.30 illustrates the principle of stereoscopic viewing.

The advantage of stereoscopic viewing over 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 measurement by photogrammetry. Not just any two images can be viewed stereoscopically, they must fulfill several conditions. The same holds for making 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, each with a slightly different meaning. A pair of images that meets the conditions of stereoscopic vision may be referred to as a stereo-image pair, a stereoscopic pair of images, stereo images, or simply as a stereo pair. A stereo pair arranged (on a computer monitor, on a table, or in a device) such 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.

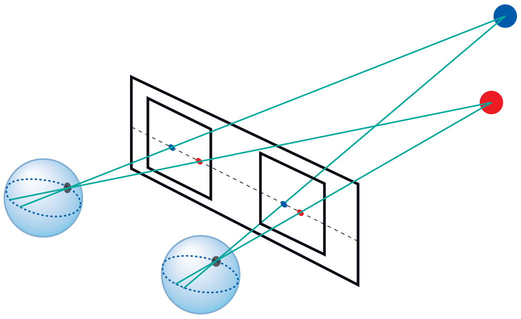

Figure 4.9
The principle of stereoscopy.

We have two options for obtaining a stereo pair with a space-borne 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 with stereo images using an airborne frame camera, we need to take strips of vertical photos/images such that the images overlap by at least 60% (see Section 4.6).

4.2 Thermal remote sensing

Thermal remote sensing is based on the measuring of electromagnetic radiation in the infrared region of the spectrum. The wavelengths most commonly used are those in the intervals 3–5 µm and 8–14 µm, in which the atmosphere is fairly transparent and the signal is only slightly attenuated by atmospheric absorption. Since the source of the radiation is the heat of the imaged surface itself (see Figures 2.6 and 2.16), 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 emitted radiation measured in the thermal wavelengths. The temperature of an object varies greatly depending on time of 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 considered 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, the one property we are interested in is, rather, how well radiation is *emitted* from the surface at different wavelengths.

- Since thermal remote sensing does not depend on reflected sunlight, it can also be done at night (for some applications this is even better than during the day).

4.2.1 Radiant and kinetic temperatures

The actual measurements by a TIR sensor will relate to the "spectral radiance" (measured in W m^{-2} sr^{-1} μm^{-1}) that reaches the sensor for a certain wavelength band. We know that the amount of radiation from an object depends on its temperature T and emissivity ϵ. That means that a cold object with high emissivity can radiate just as much radiation as a considerably hotter one with low emissivity. Often the emissivity of the object is unknown. If we assume that the emissivity of the object is equal to 1.0, then with the help of Planck's law we can calculate directly the ground temperature that is needed to create this amount of radiance in the specified wavelength band of the sensor for the object with a perfect emissivity. The temperature calculated in this way is the *radiant temperature* or T_{rad}. The terms *brightness* or "top-of-the-atmosphere" temperature are also frequently used.

radiant temperature

The radiant temperature calculated from the emitted radiation is in most cases lower than the true, *kinetic temperature* (T_{kin}) that we could measure on the ground with a contact thermometer. The reason for this 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

$$T_{rad} = \epsilon^{1/4} T_{kin}. \tag{4.3}$$

With a single thermal band (e.g. Landsat-7 ETM+), ϵ has to be estimated from other sources. One way of doing this is to do a land cover classification with all available bands and then assign an ϵ value for each class from an emissivity table (e.g. 0.99 for water, 0.85 for granite).

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 under-determined system of equations. For this reason, it is necessary to make certain assumptions about the shape of the emissivity spectrum we are trying to observe. Different algorithms exist to separate the influence of temperature from the emissivity.

4.2.2 Thermal applications

In general, applications of thermal remote sensing can be divided into two groups. In one group, the main interest is the study of surface composition by observing the surface emissivity in one or more wavelengths. In the other group, the focus is on surface temperature and its spatial and temporal distribution. The following discussion only concerns this second group.

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 4.10 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 higher contrast between the temperatures of the anomaly itself and surrounding rocks. This particular image was acquired over the Wuda coal-mining area in China in September 2002. Hotter temperatures are represented 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 (white box in Figure 4.10), several hotspots, with higher temperatures compared to the surrounding rocks, are visible. The inset shows the same mining area slightly enlarged. The hottest pixels are 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.

Glaciers monitoring With thermal remote sensing, studies of glaciers can go further than the plain observation of their extent. Understanding the dynamics of a glacier’s state requires environmental variables. Ground surface temperature is obviously among the most important variables that affect glacier dynamics.

Urban heat islands The temperature of many urban areas is significantly higher than that of surrounding natural and rural areas. This phenomenon is referred to as an urban heat island. The temperature difference is usually larger at night than during the day and occurs mainly due to the change of matter covering the land as a result of urban development: land cover in built-up areas retains heat much better than land cover in natural and rural areas. This affects the environment in many ways: it modifies rainfall patterns, wind patterns, air quality, the seasonality of vegetation growth, and so on. Urban heat islands also affect the health of urban inhabitants: in particular, they can modify the duration and magnitude of heat waves in urban areas, leading to increases in mortality rates. There are several ways to mitigate the urban heat island effect, the most prominent ones being the use of highly reflective materials and increasing the amount of urban vegetation. To study the urban heat island effect we need to observe the temperature in urban and surrounding areas. Thermal remote sensing is a suitable tool as it provides temperature measurements that incorporate the spatial extent of cities and their surroundings.

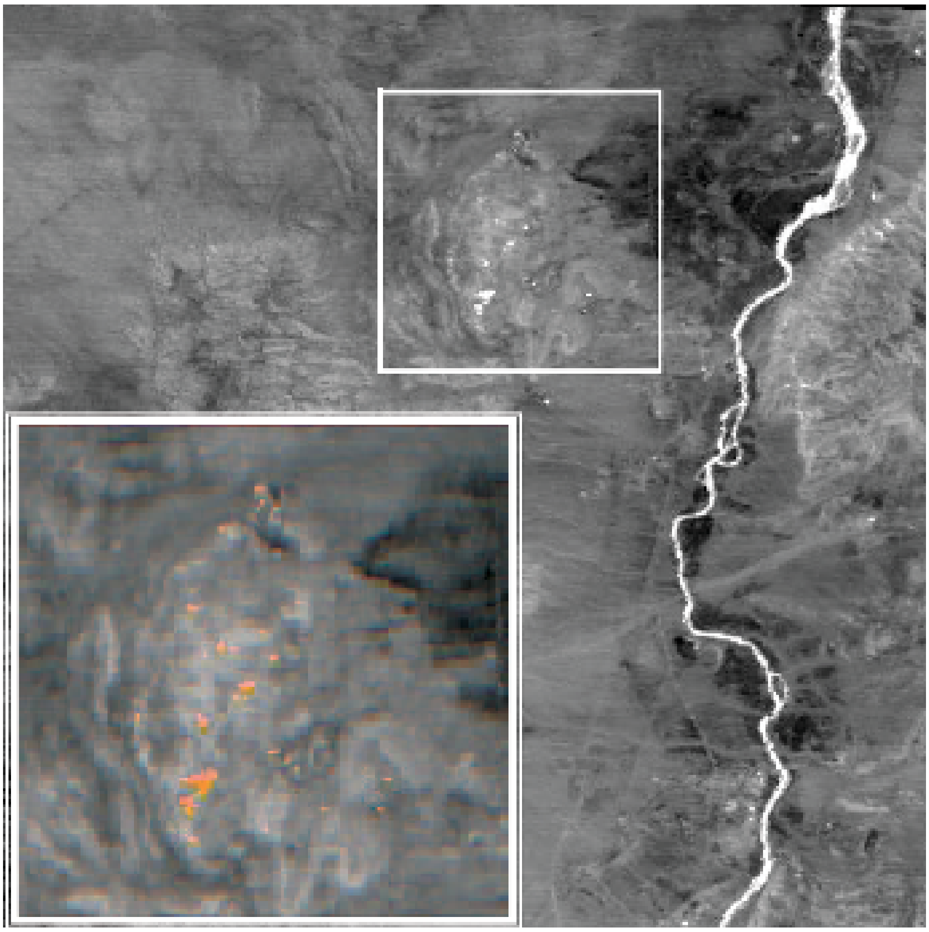

Figure 4.10
ASTER thermal band 10 over Wuda, China. Light coloured pixels inside the mining area (white box) are caused mainly by coal fires. Inset: pixels exceeding the background temperature of 18 °C are orange for better visibility of the fire locations. This scene is approximately 45 km wide.

4.3 Imaging Spectrometry

You have learnt in Section 2.5 that materials of interest may be distinguished by their spectral reflectance curves (e.g. Figure 2.14). In this section we will call spectral reflectance curves *reflectance spectra*. Most multispectral sensors that were discussed in Chapter 2 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 20–40 nm in width. Hence, broadband sensors under-sample this information and do not allow full exploitation of the spectral resolution potential available. Imaging spectrometers typically acquire images in a large number of spectral bands (more than 100). These bands are narrow (less than 10–20 nm in width) and contiguous (i.e. adjacent), which enables the extraction of reflectance spectra at pixel scale (Figure 4.11). Such narrow spectra enable the detection of diagnostic absorption features. Different names have been coined for this field of remote sensing, including imaging spectrometry, imaging spectroscopy and hyperspectral imaging.

Figure 4.12 illustrates the effect of spectral resolution for 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 reference spectrum). With each improvement in spectral resolution, the diagnostic absorption features and, therefore, the unique shape of kaolinite's spectrum become more apparent.

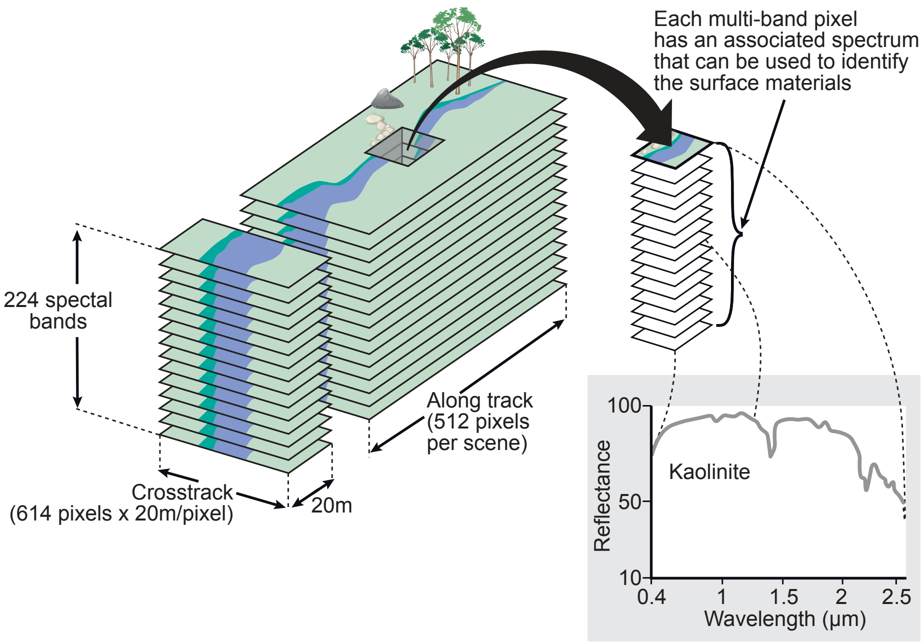

Figure 4.11
The concept of imaging spectrometry (adapted from [114]).

4.3.1 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 4.13). By studying the reflectance spectra of rocks, individual minerals and groups of minerals may be identified. In the Earth sciences, absorption in the wavelength region 0.4 µm–2.5 µm is commonly used to determine the mineralogical content of rocks. In this region, various groups of minerals have characteristic reflectance spectra; examples include phyllosilicates, carbonates, sulphates, and iron oxides and iron hydroxides. High-resolution reflectance spectra for mineralogy studies can easily be obtained in the field or the laboratory using field spectrometers.

reflectance spectra

Processes that cause absorption of electromagnetic radiation occur at the molecular and atomic levels. Two types of processes are important in the 0.4 µm–2.5 µm range: electronic processes; and vibrational processes ([21]). Depending on the molecular structure and composition, different absorption features can be identified. Reflectance spectra also correspond closely to the crystal structure of minerals and can, therefore, be used to obtain information about their crystallinity and chemical composition.

4.3.2 Pre-processing of imaging spectrometer data

Pre-processing of imaging spectrometer data involves radiometric calibration (see Section 5.2), 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 at-sensor or surface reflectance data. Section 5.2 contains an overview 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

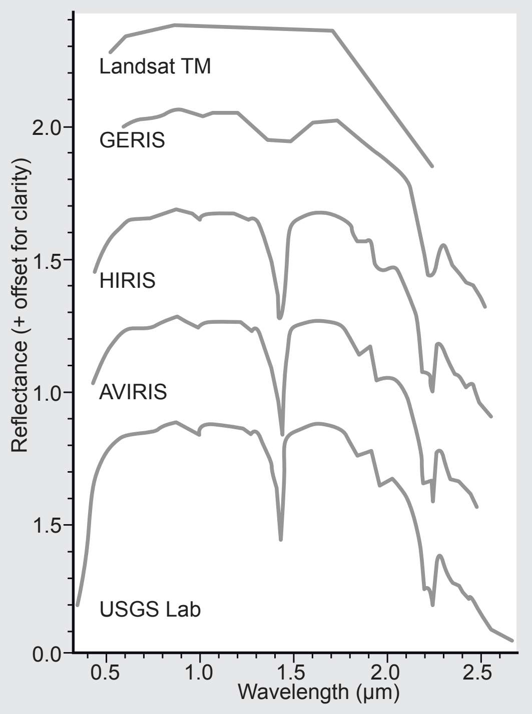

Figure 4.12
Example of a kaolinite spectrum at the original resolution (source: USGS laboratory) and at the spectral resolutions of various imaging devices. Note that the spectra are progressively offset upwards by 0.4 units for clarity (adapted from USGS).

relative correction

Alternatively, users can make a scene-dependent relative atmospheric correction using empirically derived models for the radiance-reflectance conversion that are based on calibration targets found in the imaging spectrometer data set. Empirical models often used include techniques known as flat-field correction and 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, on the other hand, requires 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 and image spectra that characterize the same ground areas, thus removing atmospheric effects, residual instrument artefacts, and viewing geometry effects.

4.3.3 Applications of imaging spectrometry data

A brief outline of current applications in various fields relevant to the thematic context of ITC are described in the remainder of this subsection.

Geology and resources exploration

Imaging spectrometry is used by the mining industry for surface mineralogy mapping, to aid in ore exploration. Other applications of this technology include lithological and structural mapping. The petroleum industry is also developing methods for using imaging spectrometry for reconnaissance surveys. The main targets are hydrocarbon seeps and microseeps.

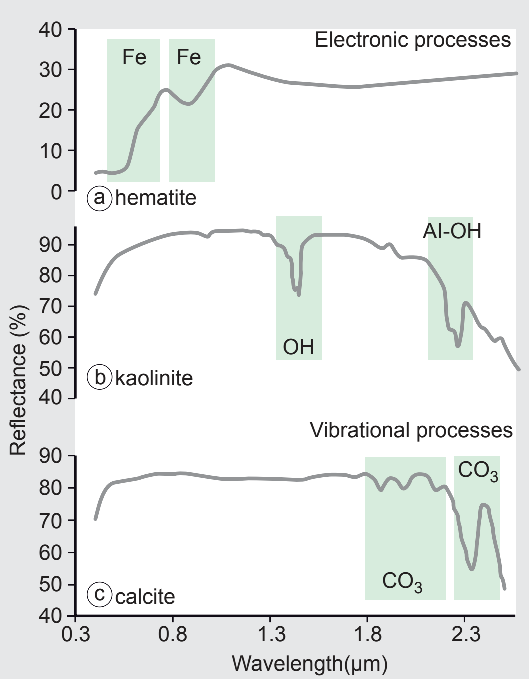

Figure 4.13
Effects of electronic and vibrational processes on absorption of electromagnetic radiation.

Other fields of application include environmental geology (and related geo-botany), in which currently much work is being done on acid mine drainage and mine-waste monitoring. Imaging of the atmospheric effects resulting from geological processes (e.g. sulfates emitted from volcanoes), to predict and quantify the presence of various gases for hazard assessment, is also an important field. In soil science, much emphasis has been placed on the use of spectrometry for the study of soil surface properties and soil composition analysis. Major elements such as iron and calcium, in addition to cation–anion 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-nature areas. Recently, spectral identification from imaging spectrometers has been successfully applied to the mapping of the swelling clay 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 already using imaging spectrometer data for reconnaissance-level exploration.

Vegetation sciences

Much research in vegetation studies has emphasized leaf biochemistry and leaf and canopy structure. Biophysical models for leaf constituents are currently available, as are soil–vegetation models. Estimates of plant material and structure, and biophysical variables, include carbon balance, yield/volume, nitrogen, cellulose, and chlorophyll. Leaf area index and vegetation indices have been extended to the hyperspectral domain and remain important physical variables for characterizing vegetation. One ultimate goal is the estimation of biomass and the monitoring of changes therein. Several research groups have been investigating the bi-directional reflectance function in relation to vegetation species analysis and floristics. Vegetation stress as a result of water deficiency, pollution (such as acid mine drainage) and geo-botanical anomalies in relation to ore deposits or petroleum and gas seepage links vegetation analysis to exploration. Another upcoming field of application is precision agriculture, in which imaging spectrometry is being used to improve agricultural practices. An important factor in the health of vegetation is chlorophyll absorption and, in relation to that, the position of the red edge, determined using the red-edge index. Red edge is the name given to the steep increase in the reflectance spectrum of vegetation between visible red and near infrared wavelengths.

Hydrology

In hydrological sciences, the interaction of electromagnetic radiation with water, and the inherent and apparent optical properties of water are a central issue. Atmospheric correction and air–water interface corrections are very important in the imaging spectrometry of water bodies. Water quality of freshwater aquatic environments, estuarine environments and coastal zones usually has an important impact on national water bodies. Detection and identification of phytoplankton biomass, suspended sediments and other matter, coloured dissolved organic matter, and aquatic vegetation (i.e. macrophytes) are crucial variables in optical models of water quality. Much emphasis has been put on the mapping and monitoring of the state and growth or breaking down of coral reefs, as these are important for the CO_2 cycle. In general, many multisensor missions such as Terra and Envisat are directed towards integrated approaches for global climate change studies and global oceanography. Atmospheric models are important in global climate-change studies and aid in the correctin of optical data for scattering and absorption owing to trace gases in the atmosphere. In particular, the optical properties and absorption characteristics of ozone, oxygen, water vapour and other trace gases, and scattering by molecules and aerosols, are important variables in atmosphere studies. All these can be and are estimated from imaging spectrometry data.

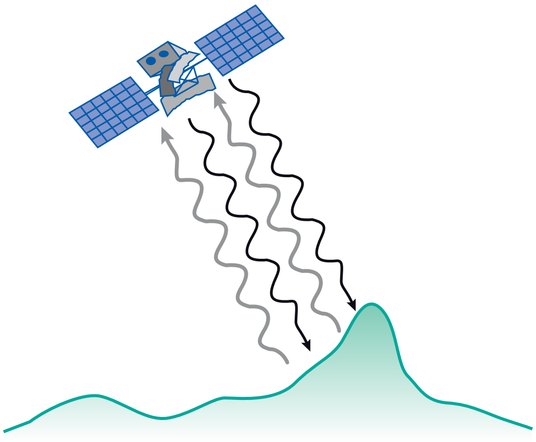

Figure 4.14
Principle of active microwave remote sensing.

4.4 Radar

4.4.1 What is radar?

microwave RS

Microwave remote sensing uses electromagnetic waves with wavelengths between 1 cm and 1 m (Figure 2.5). These relatively long wavelengths have the advantage that they can penetrate clouds and are not affected by atmospheric scattering. Although microwave remote sensing is primarily considered to be an active technique, passive sensors are also used. Microwave radiometers operate, similarly to thermal sensors, by detecting naturally emitted microwave radiation (either terrestrial or atmospheric). They are primarily used in meteorology, hydrology and oceanography.

In active systems, the antenna emits microwave signals to the Earth's surface, where they are backscattered. The part of the electromagnetic radiation that is scattered back in the direction of the antenna is detected by a sensor, as illustrated in Figure 4.14. There are several advantages to be gained from using active sensors, which have their own source of EM radiation:

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

non-imaging radar

Active sensors can be divided into two types: 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 component and *ranging* is another term for distance. Radar sensors were originally developed and used by the military. Nowadays, radar sensors are also widely used in civilian applications, such as environmental monitoring. Examples of non-imaging microwave instruments are *altimeters*, which collect distance information (e.g. sea-surface elevation), and *scatterometers*, which acquire information about object properties (e.g. wind speed).

The following subsection 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 the physical interaction of the waves with the Earth's surface. The interactions that take place and how radar images can be interpreted are also explained.

4.4.2 Principles of imaging radar

Imaging radar systems have a number of 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 reaching the antenna and filters and amplifies it as required for recording. The recorder then stores the received signal.

backscattered radiation

Imaging radar acquires an image in which each pixel contains a digital number according to the strength of the backscattered radiation received from the ground. The radiation received from each emitted radar pulse can be expressed in terms of the physical variables and illumination geometry according to the *radar equation*:

$$P_r = \frac{G^2 \lambda^2 P_t \sigma}{(4\pi)^3 R^4}, \tag{4.4}$$

where

P_r = received radiance,
G = antenna gain,
λ = wavelength,
P_t = emitted radiance,
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.

This equation demonstrates that there are three main factors that influence the strength of the backscattered radiation received:

- radar system properties, i.e. wavelength, antenna and emitted power;
- radar imaging geometry, which defines the size of the illuminated area, which is in turn a function of, for example, beam width, incidence angle and range;
- the characteristics of interaction of the radar signal with objects, i.e. surface roughness and composition, and terrain relief (magnitude and orientation of slopes).

  These factors are explained in more detail below.

What exactly does a radar system measure? To interpret radar images correctly, it is important to understand what a radar sensor detects. Radar waves have the same physical properties as those explained in Chapter 2. Radar waves, too, have electric and magnetic fields that oscillate as a sine wave in perpendicular planes. In dealing with radar, the concepts of wavelength, period, frequency, amplitude, and phase are therefore relevant.

A radar transmitter creates microwave signals, i.e. *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 back to the antenna, where the signal is received and its intensity measured. The signal needs to travel twice the distance between an object and the receiver/antenna. As we know

intensity

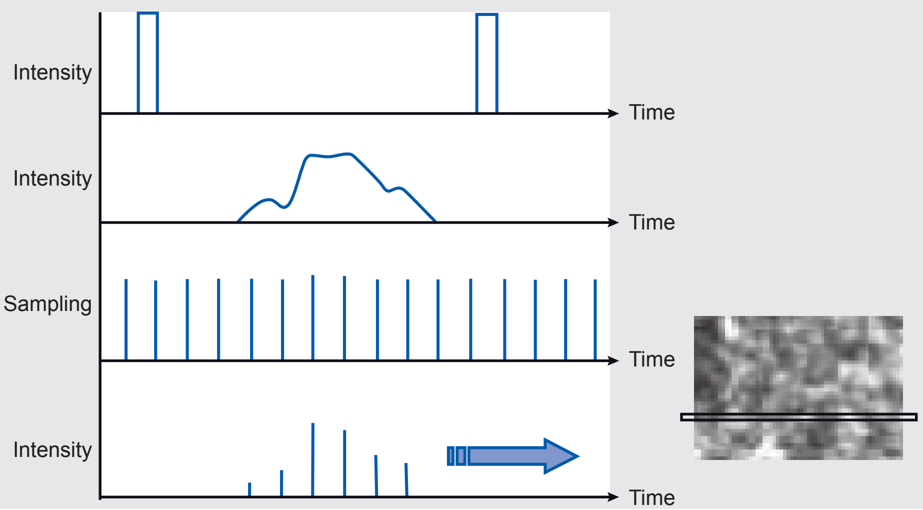

Figure 4.15
Illustration of how radar pixels result from pulses. For each sequence shown, one image line is generated.

the speed of light, we can calculate the distance (*range*) between sensor and object (see Formula 4.5).

To create an *image*, the return signal of each single pulse is sampled and samples stored in an image line (Figure 4.15). 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 within one or more different bands. For better identification, a standard has been established that defines various wavelength ranges using letters to distinguish them from each other (Figure 4.16); you can recognize the different wavelengths used in radar missions from the letters used. The European ERS mission and the Canadian Radarsat use, for example, C-band radar. Just like multispectral bands, different radar bands provide information about different object characteristics.

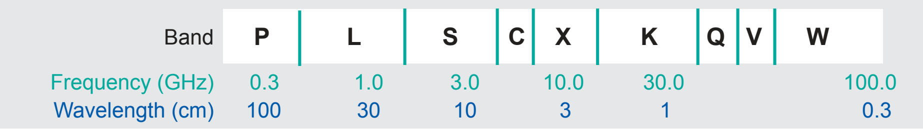

Figure 4.16
The microwave spectrum and band identification by letters.

Microwave polarizations The polarization of an electromagnetic wave is important in radar remote sensing. Depending on the orientation of the emitted and received radar wave, polarization will result in different images (see Figure 2.1, 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, e.g. to classify agricultural fields. In radar system descriptions you will come across the following abbreviations:

- HH: horizontal transmission and horizontal reception;
- VV: vertical transmission and vertical reception;

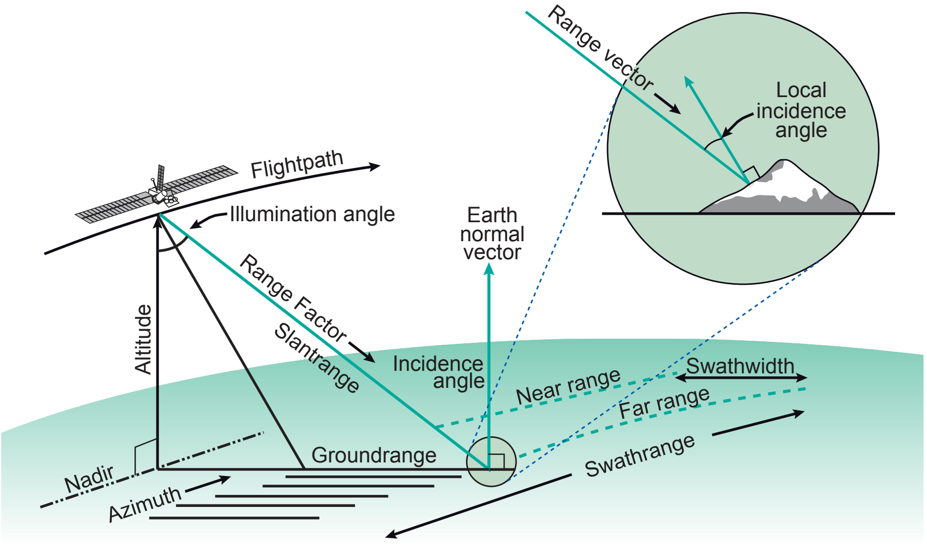

Figure 4.17
Radar remote-sensing geometry.

- HV: horizontal transmission and vertical reception;
- VH: vertical transmission and horizontal reception.

4.4.3 Geometric properties of radar

The platform carrying the radar sensor travels along its orbit or flight path (Figure 4.17). 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 nadir, i.e. side-looking. The direction along-track is called *azimuth*and the direction perpendicular (across-track) is 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 nadir is called *far range* (Figure 4.17). The *incidence angle* of the system is defined as the angle between the radar beam and the local Earth normal vector. 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 the curvature of the Earth (Figure 4.17). The local incidence angle is defined as the angle between the radar beam and the local surface normal vector. The radar sensor measures the distance between antenna and object. This line is called the *slant range*. The true horizontal distance along the ground corresponding to each point of measured slant range is called the *ground range* (Figure 4.17).

ranges

Spatial resolution

In radar remote sensing, the images are created from the backscattered portion of emitted signals. Without further sophisticated processing, the spatial resolutions of 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). As different parameters determine the spatial resolution in range and azimuth, it is obvious that the spatial

RAR

resolution in each direction is different from the other. For radar image processing and interpretation it is useful to resample the data to the same GSD in both directions.

Slant range resolution For 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. Slant range resolution is independent of the actual range (see Figure 4.18).

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

aperture

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 capability of penetrating clouds. To improve the spatial resolution, a large antenna is synthesized 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 flight path in which the object is "visible". Most airborne and space-borne radar systems use this type of radar. Systems using this approach are referred to as *Synthetic Aperture Radar (SAR)*.

SAR

4.4.4 Distortions in radar images

Due to the side-looking 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; see Figure 4.19). Interference due to the coherence of the signal causes *speckle* effects.

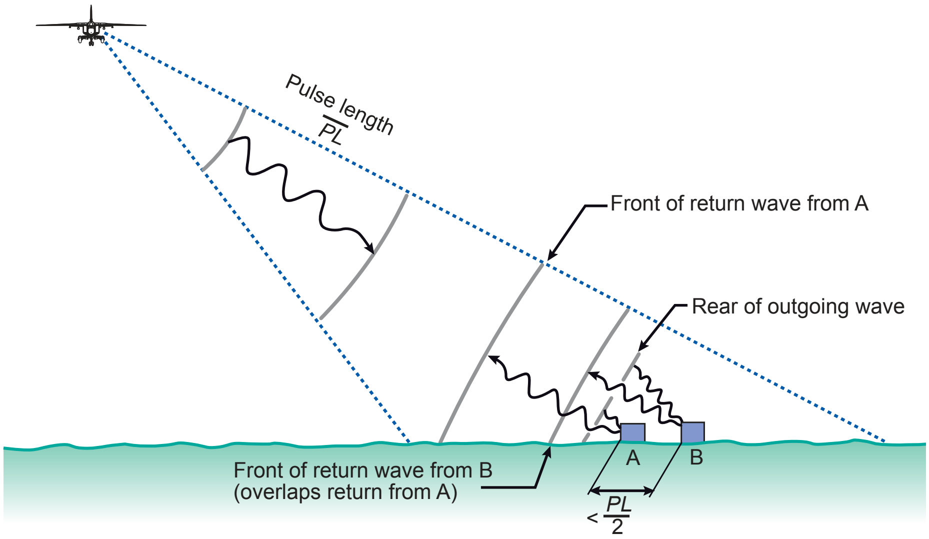

Figure 4.18
Illustration of the slant range resolution.

Scale distortions

Radar measures ranges to objects in slant range rather than true horizontal distances along the ground. Therefore the image has different scales moving from near to far range (Figure 4.17). This means that objects in near range are compressed as compared to objects in far range. For proper interpretation, the image has to be corrected and transformed into ground range geometry.

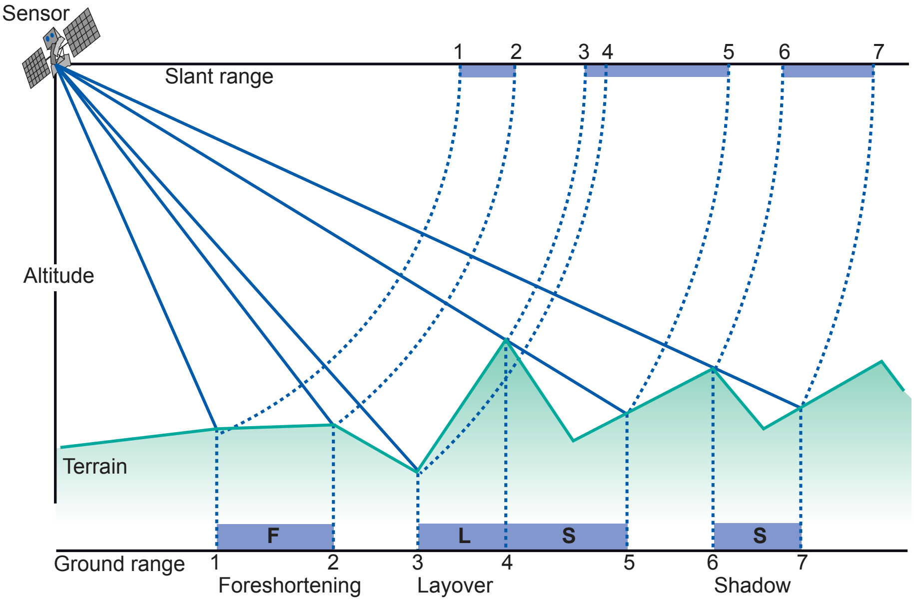

Figure 4.19
Geometric distortions in a radar image caused by varying terrain elevation.

Terrain-induced distortions

Similarly to optical sensors that can operate in an oblique manner (e.g. 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* (see Figure 4.19).

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, i.e. the slope “lays over”. As you can understand from the definition of foreshortening, layover is an extreme case of foreshortening. Layover areas in the image are very bright.

shadow

In the case of slopes that are facing away from the sensor, the radar beam cannot illuminate the area at all. Therefore, there is no radiation that can be backscattered to the sensor and so those regions remain dark in the image.

Radiometric distortions

Geometric distortions also influence the received radiation. Since backscattered radiation is collected in slant range, the received radiation coming from a slope facing the sensor is stored in a reduced area in the image, i.e. 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 radiation collected from different objects is combined. Slopes facing the radar appear 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 appearance of the image and therefore contribute to an understanding of surface structure and terrain relief.

speckle

interference

A typical property of radar images is *speckle*, which appears as grainy "salt and pepper" effects in the image (Figure 4.20). Speckle is caused by the interference of backscattered signals coming from an area that is encapsulated 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 homogeneous area. Speckle degrades the quality of the image and makes the interpretation of radar images difficult.

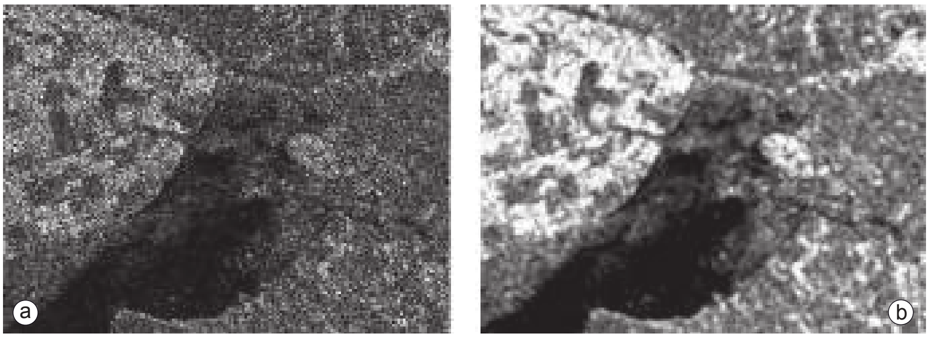

Figure 4.20
An original (a) and speckle filtered (b) radar image.

Speckle reduction

It is possible to reduce speckle by 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 to the images. Speckle filters are designed to adapt to local image variations in order to smooth values, thus reducing speckle and enhancing lines and edges to maintain the sharpness of an image.

4.4.5 Interpretation of radar images

The brightness of features in a radar image depends on the strength of the backscattered signal. In turn, the amount of radiation that is backscattered depends on a number of factors. An understanding of these factors helps with the proper interpretation of radar images.

Microwave signal and object interactions

For those who are concerned with the visual interpretation of radar images, the degree to which they are able to interpret an image depends upon whether they can identify typical/representative tones related to surface characteristics. The amount of radiation that is received at the radar antenna depends on the illuminating signal (radar system variables such as wavelength, polarization and viewing geometry) and the characteristics of the illuminated object (e.g. roughness, shape, orientation, dielectric constant).

Surface roughness is the terrain property that most strongly influences the strength of radar backscatter. It is determined by textural features comparable to the size of the radar wavelength (typically between 5 and 40 cm), for example, leaves and twigs of vegetation and sand, gravel and cobble stones. 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 radiation away from the antenna without returning a signal, thereby resulting in a black image. With an increase in surface roughness, the amount of radiation 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 radiation returned, the brighter the signal is shown on 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, which 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 [112]. 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 the 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 variations 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 radiation will be backscattered from the vegetation 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, as well as the characteristics of the radar used, such as wavelength and its related effective penetration depth [8].

Point objects are objects of limited size that give a very strong radar return signal. Usually, the high level of backscatter is caused by *corner reflection*. An example of this is the dihedral corner reflector—a point object situation resulting from two flat surfaces intersecting at 90° and situated orthogonally to the incident radar beam. Common forms of dihedral configurations are man-made features such as transmission towers, railway 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

corner reflection

three mutually perpendicular flat surfaces. Point objects that are corner reflectors are commonly used to identify known fixed points in an area in order to perform precise calibration measurements. Such objects can occur 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 a radar image and are typical for urban areas. Point objects are examples of objects that are sometimes below the resolution of a radar system but, because they dominate the return signal from a cell, nevertheless give a clearly visible point; they may even dominate the surrounding cells.

4.4.6 Applications of radar

There are many useful applications of radar imaging. Radar data provide information complementary 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 can also be used to distinguish between different types of land cover, e.g. urban areas, agricultural fields and water bodies. In urban areas, radar detects buildings (corner reflectors) and metal constructions, thus allowing the extent of urban areas to be delineated, which is a key observable for urban growth studies. 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 moment (season) to obtain the necessary parameters. This is possible because radar can operate independently of weather or light 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. Since radar backscatter is sensitive to surface roughness, it helps to discriminate between ice and debris, thus making it potentially suitable for glaciers monitoring studies. Radar also allows the measurement of elevation and change in elevation by a technique called Interferometric SAR (INSAR). Radar has also been successfully applied in 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.

4.5 Laser scanning

4.5.1 Basic principles

measuring by three sensors

Laser scanning—in functional terms—can be defined as a system that produces digital surface models. The system comprises an assembly of various sensors, recording devices and software. The core component is the laser mechanism. The laser measures distance, which is referred to as "laser ranging". When mounted on an aircraft, a laser rangefinder measures at very short time intervals the distance to the terrain. -Combining a laser rangefinder with sensors that can measure the position and attitude of the aircraft (GPS & IMU) makes it possible to create a model of the terrain surface in terms of a set of (X, Y, Z) coordinates, following the polar measuring principle; see Figure 4.21.

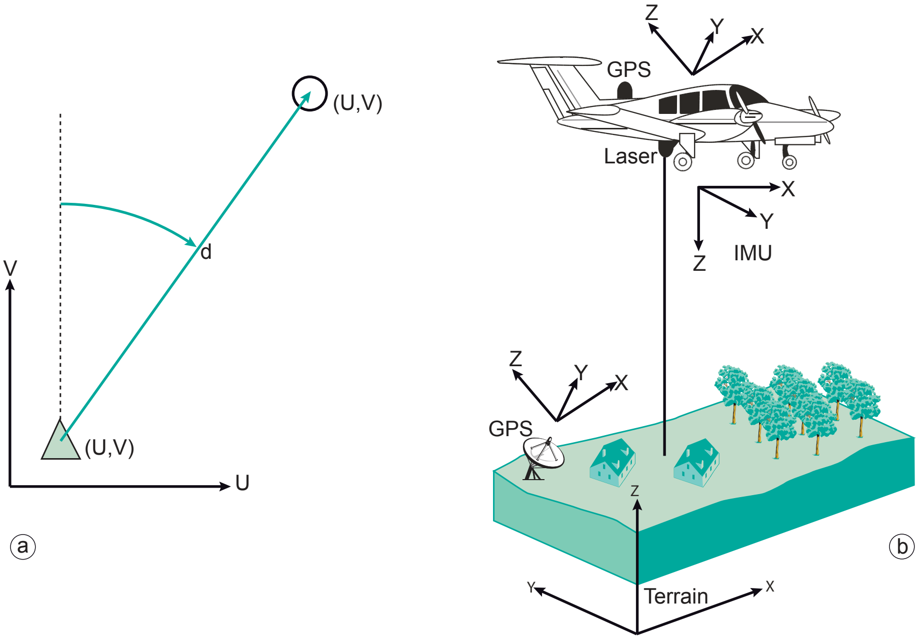

Figure 4.21
Polar measuring principle (a) and its application to ALS (b).

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), i.e. we model the surface of interest by providing its elevation at many points, each 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 4.22). Post-processing is required to obtain a digital terrain model (DTM) from the DSM.

applications of ALS

The key advantages of ALS are its high ranging precision, its ability to yield high resolution DSMs in near real-time, and its complete or nearly complete independence of weather, season or light conditions. Typical applications of ALS are, therefore, forest surveys, surveying of coastal areas and sand deserts, flood-plain mapping, power-line and pipeline mapping, monitoring open-pit mining and 3D city modelling.

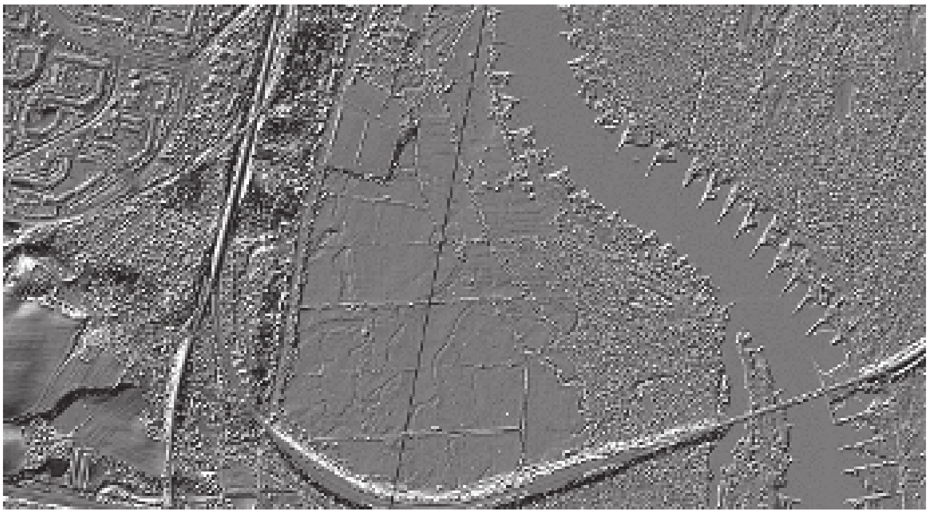

Figure 4.22
DSM of part of Frankfurt (Oder), Germany (1 m point spacing). Courtesy of TopoSys.

4.5.2 ALS components and processes

laser

LASER stands for Light Amplification by Stimulated Emission of Radiation. Although he did not invent it, Einstein can be considered the father of the laser. Roughly 85 years ago he postulated the phenomena of photons and stimulated emission; he won the Nobel prize for related research on the photoelectric effect. In 1960, Theodore Maiman, employed 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. Also specific for lasers is the very high intensity of radiation they emit. Today, lasers are used for many purposes, even human surgery. Lasers can damage cells (by boiling their water content), so they are a potential hazard for the eye. Users of laser rangefinders therefore have to attend safety classes; safety rules must be strictly observed when using lasers for surveying applications.

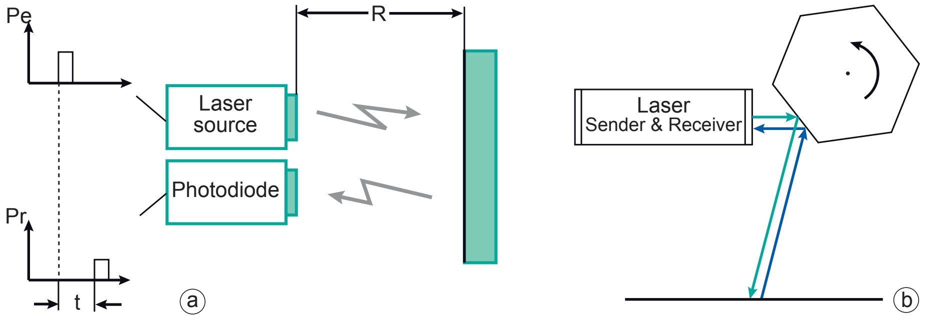

Figure 4.23
Concept of laser ranging and scanning [127].

laser rangefinders

Laser rangefinders and scanners come in various forms. Most airborne laser instruments are "pulse lasers". A pulse is a signal of very short duration that travels as a beam. Airborne laser rangefinders for topographic applications emit NIR radiation. An emitted pulse is reflected by the ground and its return signal is sensed by a photodiode (Figure 4.23). A time counter starts when a pulse is sent out and stops on its return. The elapse time is measured with a resolution of 0.1 ns. As we know the speed of light,c, the elapsed time can easily be converted to a distance:

$$R = \frac{1}{2}\,ct. \tag{4.5}$$

Modern laser scanners send out pulses at a very high frequency (up to 300,000 pulses per second). Across-track scanning is in most cases achieved by a moving mirror, which deflects the laser beam. 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 (a swath of points) can be captured on a single leg of , instead of just a line of points as was the case with earlier versions of laser systems (laser profilers).

laser scanner

Simple laser rangefinders register one return pulse for every emitted pulse. By contrast, modern laser rangefinders for airborne applications record multiple echoes from the same pulse. Multiple-return laser ranging is specifically relevant for aerial surveys of terrain covered by vegetation because it helps distinguish vegetation echoes from ground echoes. For a pulse that hits a leaf at the top of a tree, part of the pulse may be reflected, while another part of it may travel further, perhaps hitting a branch and, eventually, even the ground; see Figure 4.24. Many of the first return "echoes" will be from the tree canopy, while the last returns are more likely to come from the ground. Each return can be converted to an $(X,\ Y,\ Z)$ of the illuminated target point. To figure out whether the point is on the ground or somewhere amongst the vegetation is far from trivial. Multiple return ranging does not give a direct answer but it helps find one. An example of a first return and last return DSM is shown in Figure 4.25. *Full waveform sensors* represent the further development of this approach. Instead of only detecting an echo if its intensity is above a certain threshold (Figure 4.24), full waveform scanners or altimeters (as on ICESat, see below) digitize the entire return signal of each emitted laser pulse. Full waveform laser rangefinders can provide information about surface roughness and more cues on vegetation cover.

multiple return ranging

full waveform sensors

As well as measuring the range, some laser-based instruments also measure the amplitude of the reflected signal to obtain an image (often referred to as "intensity image"). Imaging by laser scanner is different from imaging by radar instruments. 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 images obtained 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 necessary elements. Measuring the distance from the aircraft to the terrain can be done very precisely by the laser rangefinder (accurate to within centimetres), and we can accurately determine the position and altitude of the aircraft using a POS (Section 4.1.1).

GPS and IMU

The most widely used platforms for ALS are airplanes and helicopters. Helicopters are better suited for very high-resolution surveys, because they can easily fly slowly. The minimum flying height is, among other things, dependent on the safe eye–laser distance for the instrument. The major limiting factor of the maximum flying height is energy loss of the laser beam. 1000 m and less are frequently used flying altitudes, although there are systems for which heights of 8000 m are feasible.

ALS platforms

Unlike aerial surveys for generating stereo coverage of photographs—for which each terrain point should be recorded at least twice—in ALS a terrain point is, in principle, only "collected" once, even if the strips flown overlap. This is an advantage when surveying urban areas and forests, but it has disadvantages for error detection.

After the flight, the recordings from the laser instrument and the POS are co-registered co-registering the data

to the same time and then converted to (X, Y, Z) coordinates 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 (e.g. to compute timber volumes) and, in particular, filtering 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 “devegging” for short, a term left over from the early days when ALS was primarily used for forested areas (Figure 4.26).

Proper system calibration, accurate flight planning and execution (including the GPS logistics), and adequate software are critical factors in ensuring one gets the right data at the right time.

4.5.3 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 then, why go for ALS? There are in fact several good reasons for using ALS for terrain modelling:

- A laser rangefinder measures distance by recoding the elapse time between emitting a pulse and receiving the reflected pulse from the terrain. Hence, the laser rangefinder is an active sensor and can be used both during daylight hours

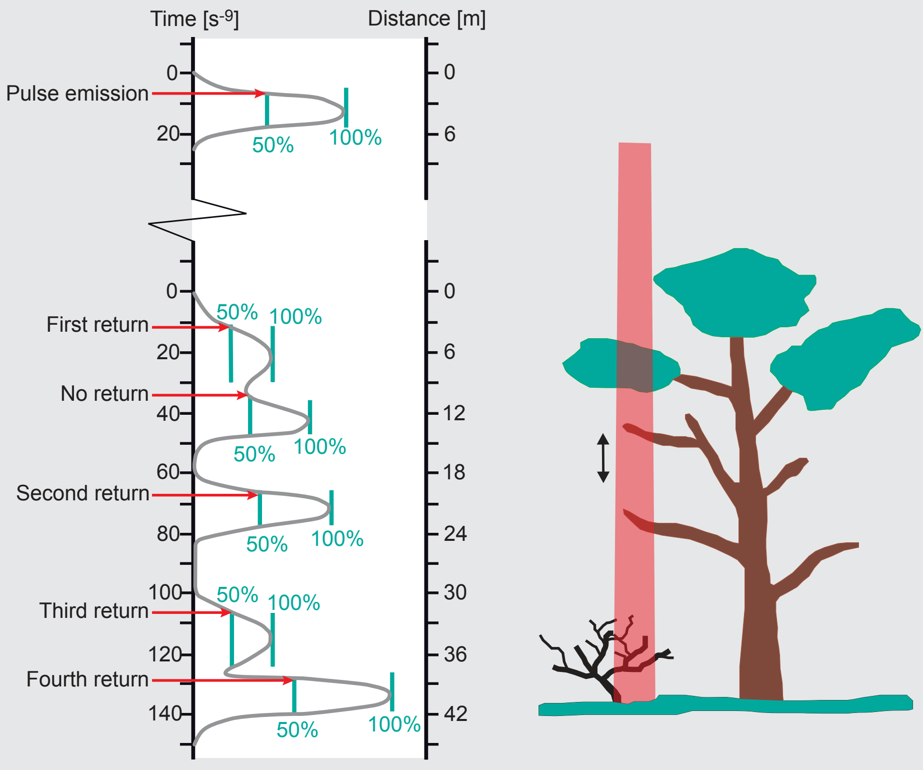

Figure 4.24
Multiple-return laser ranging. Adapted from Mosaic Mapping Systems, Inc.

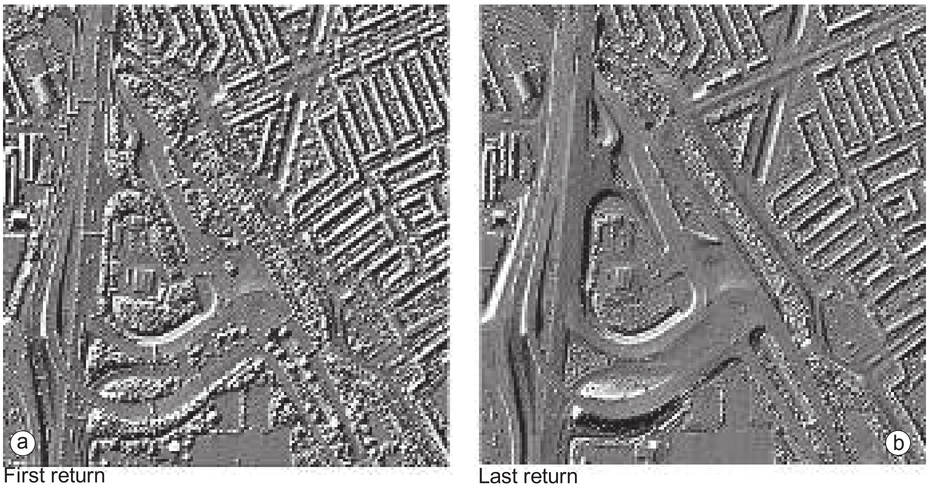

Figure 4.25
First (a) and last (b) return DSMs of the same area. Courtesy of TopoSys.

and at night. The possibility of flying at night comes in handy when, for instance, surveying a busy airport.

- Unlike indirect distance measuring done using stereo images, laser ranging does not depend on surface/terrain texture.
- Laser ranging is less weather-dependent than passive optical sensors. A laser cannot penetrate clouds as microwave radar can, but it can be used at low altitudes, 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 can, therefore, have a diameter smaller than 20 cm (depending on the laser type and flying height). The simplifying assumption of "measuring points" is thus closely approximated. ALS can "see" objects that are much smaller than the footprint of the laser beam, making it suitable for mapping power lines.
- A laser beam cannot penetrate leaves, but it can pass through the tree canopy, unless that is very dense.
- A laser rangefinder can measure distances very precisely and very frequently; therefore a DSM with a high density of points can be obtained with accurate elevation values. The attainable elevation (vertical coordinate) accuracy with 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 attractive topics for researchers.
- The entire data collection process is digital, which allows it to be automated to a high degree, thus facilitating fast processing.
- Other than a calibration site, which can usually be set up near the airfield, ALS does not need any ground control.

There are two additional major advantages of laser ranging compared to microwave radar: high frequency X pulses can be generated at short intervals and highly directional beams can be emitted. The latter is possible because of the short wavelength of

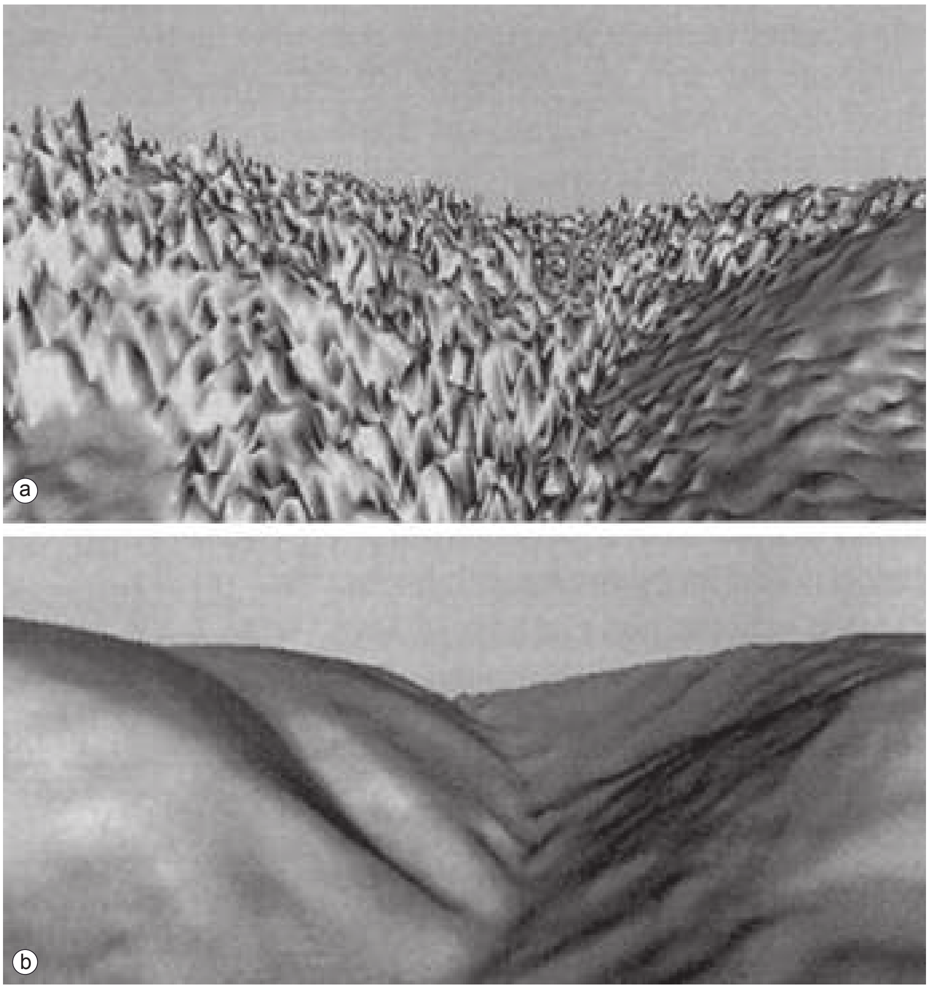

Figure 4.26
Devegging laser data: filtering a DSM (a) to create a DTM (b). From [61].

lasers (10,000 to 1,000,000 times shorter than microwaves). The consequence is much higher ranging accuracy.

Note that the term *radar* is often used as a short form for microwave radar. In the literature, however, you may also come across the term "laser radar", which is synonymous for laser ranging. A more frequently used synonym for laser ranging is LIDAR, although there are also LIDAR instruments that do not measure the distance to but, rather, the velocity of a target ('Doppler LIDARs').

Glacier monitoring Laser scanning provides information on surface elevation, thus making it a potentially useful tool for monitoring glaciers. However, modern laser scanners usually provide information at very fine spatial resolutions, which are not required in glacier studies. Furthermore, to differentiate glacier ice from debris, we require additional information from other sources.

Urban growth The sensitivity of laser scanning to the geometric properties of surfaces makes it a suitable tool for detecting objects of urban infrastructure. Laser scanning is, therefore, a potentially useful tool for urbanization studies, especially when very detailed spatial information is required, such as for detecting informal settlements and city construction works. Detailed information about terrain is also relevant for the modelling and monitoring of city growth. Nevertheless, this technique is not currently being used in urban growth studies.

4.6 Aerial photography

Introduction

Aerial photographs have been used since the early 20th century to provide geospatial data for a wide range of applications. *Photography* is the process or art of producing images by directing light onto a light-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 that are meaningful in the terrain. Some of ITC's early activities included photography and photogrammetry, the latter being, at that time, the most innovative and promising technique available for the topographic mapping of large areas. Aerial film cameras are typically mounted on aircraft, although a Russian satellite is known to have carried a photographic camera and NASA Space Shuttle missions have systematically photographed all aspects of their flights.

Aerial photographs and their digital variant, obtained by digital frame cameras, are today the prime data source for medium- to large-scale topographic mapping and for many cadastral surveys and civil engineering projects, as well as 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 over the course of many years, so new, significant developments cannot be expected. Owners of aerial film cameras will continue to use them as long as Agfa and Kodak continue to produce film at affordable prices.

vertical photo

Two broad categories of aerial photographs can be distinguished: *vertical* and *oblique* photographs (Figure 4.27). For most mapping applications, vertical aerial photographs are required. A vertical aerial photograph is produced with a camera mounted into the floor of a survey aircraft. The resulting image is similar to a map and has a scale that is roughly 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 measurement.

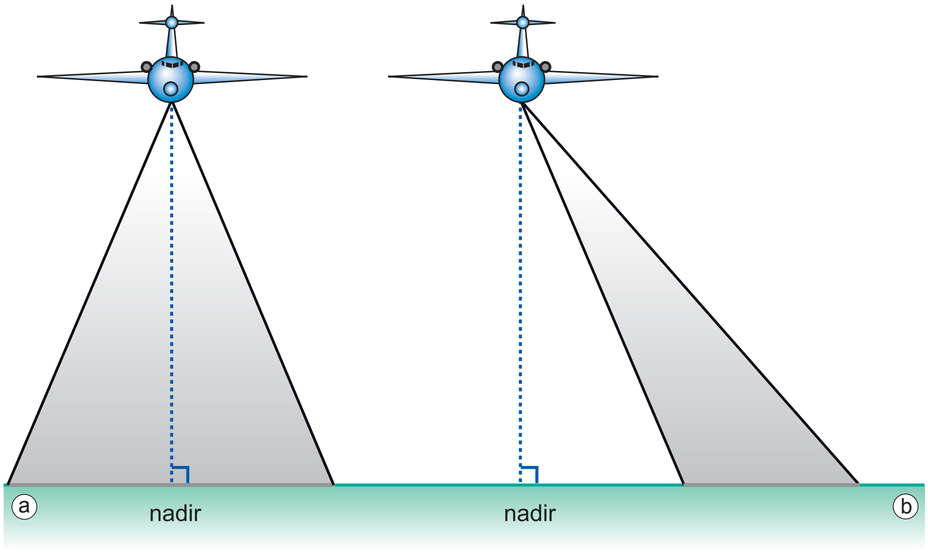

Figure 4.27
Vertical (a) and oblique (b) aerial photography.

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

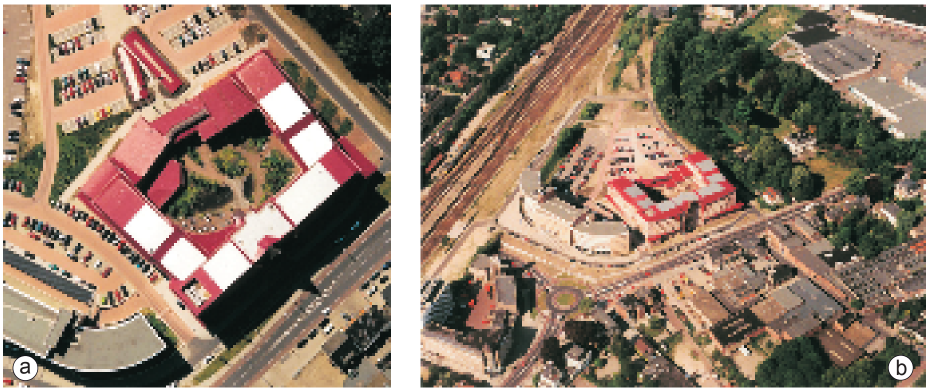

Figure 4.28
A vertical (a) and oblique (b) aerial photograph of the ITC building, 1999.

also be made using a hand-held camera and shooting through an open window of an aircraft. The scale of an oblique photo varies from the foreground to the background, which complicates the measurement of positions from the image. For this reason, oblique photographs are rarely used for mapping. Nevertheless, oblique images can be useful for obtaining side views of objects such as buildings.

oblique photo

This section discusses the aerial photo camera, films and methods used for vertical aerial photography. Subsection 4.6.1 describes the aerial camera and its main components. In broad terms, photography is based on the exposure of a photographic film to light, the processing of the film, and the printing of photographs from the processed film. Subsection 4.6.2 discusses the basic geometric—i.e. spatial—characteristics of aerial photographs. Finally, the basics concepts of aerial photography missions are introduced in Subsection 4.6.3.

4.6.1 Aerial survey cameras

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; their cameras are the RC-30 and the RMK-TOP, respectively. Aerial survey cameras contain a number of components that are also common to any typical hand-held camera, as well as a number of specialized components that are necessary for its specific role. The large size of aerial cameras results from the need to acquire images of large areas with a high spatial resolution. This is achieved by using very large-sized film. Modern aerial survey cameras produce negatives measuring 23 cm × 23 cm (9 inch × 9 inch); up to 600 photographs may be recorded on a single roll of film. To achieve the same degree of quality as an aerial film camera, a digital camera has to produce shots comprising about 200 million pixels.

4.6.2 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 available in different focal lengths.

Scale

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

$$s = \frac{H}{f}. \tag{4.6}$$

Obviously, the same scale can be achieved with different combinations of focal length and flying height. If a lens of smaller focal length is used, while the flying height remains constant, then (see also Figure 4.29):

- The *photo scale factor* will increase and the size of individual details in the image will become smaller. In the example shown in Figure 4.29, using a 150 mm and 300 mm lens at H =2000 m results in scale factors of 13,333 and 6,666, respectively.
- The ground coverage increases. A 23 cm × 23 cm negative covers an area of 3066 m × 3066 m if $f = 150$ mm. The width of the coverage reduces to 1533 m if $f = 300$ mm. Subsequent processing takes less time if we can cover a large area with fewer photos.
- The angular field of view increases and the image perspective changes. The FOV for a wide-angle lens is 74°; for a normal angle lens (300 mm) it is 41°. Using shorter focal lengths has the advantage of giving more precise elevation measurements in stereo images (see Section 5.3). Flying a camera with a wide-angle lens at low altitudes has the disadvantage of producing larger obscured areas: if there are tall buildings near the edges of a photographed scene, the areas behind the buildings become hidden because of the central perspective; we call this the *dead ground effect*.

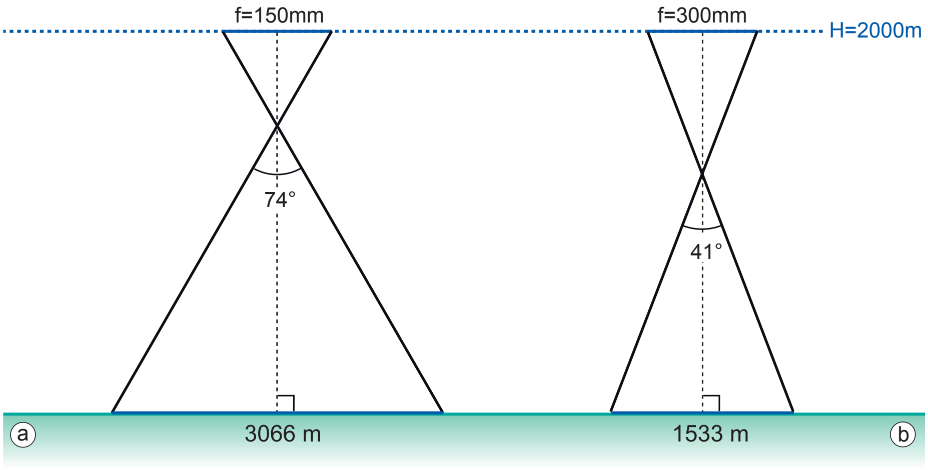

Figure 4.29
The effect of different focal lengths on ground coverage for the same flying height.

Spatial resolution

While scale is a generally understood and applied term, the use of *spatial resolution* in aerial photography is quite difficult. *Spatial resolution* refers to the ability to 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:

- the image scale factor—spatial resolution decreases as the scale factor increases;

- the quality of the optical system—expensive high-quality aerial lenses perform much better than the inexpensive lenses in 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 this list we can conclude that the actual value of resolution for an aerial photograph depends on quite a number of factors. The most variable factor is the atmospheric conditions, which can change from mission to mission and even during a mission.

4.6.3 Aerial photography missions

Mission planning When a mapping project requires aerial photographs, some of the first tasks to be done are to select the required photo scale factor, the type of lens to be used, the type of film to be used, and the required percentage of overlap (for stereo pairs). Forward overlap is usually around 60%, while sideways overlap is typically around 20%; Figure 4.30 shows a survey area covered by a number of flight lines. In addition, the date and time of acquisition should be considered with respect to growing season, light conditions and shadow effects.

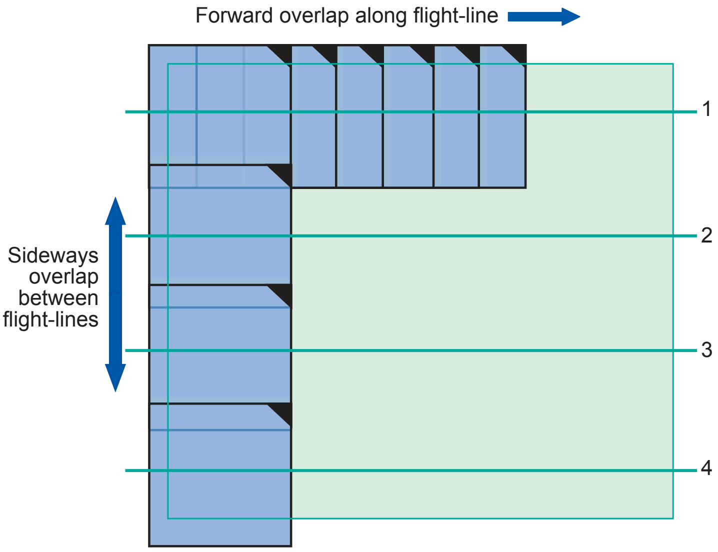

Figure 4.30
Arrangement of photos in a typical *aerial photo block*.

Once the required scale is defined, the following parameters can be determined:

- the required flying height,
- 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 fed into a mission guidance system.

Mission execution In current professional practice, 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, the number of images required, and constraints regarding time of day or Sun angle, season, and atmospheric conditions.

With the camera positions loaded into a mission guidance system, the pilot is then guided—with the support of GPS—along the mission's flight lines such that deviation from the ideal line (horizontal and vertical) and time to the next exposure station is shown on a display (together with other relevant parameters). If the aircraft passes "close enough" to a predetermined exposure station, the camera fires automatically at the nearest position. This makes it possible to have the data of several projects on board, so that the pilot can choose a project (or part of a project) according to prevailing local weather conditions. If necessary, one can also abandon a project and resume it later.

In the absence of GPS guidance, the aircraft's navigator has to observe the terrain using the traditional viewing system of the aerial camera, check actual flight lines against the planned ones, which are shown graphically on topographic maps, give the required corrections (e.g. to the left or to the right) to the pilot, and tune the overlap regulator to the apparent forward speed of the airplane.

Satellite-based positioning systems and IMU provide a means for achieving accurate navigation. They offer precise positioning of the aircraft during a mission, ensuring that the photographs are taken at the correct points. Computer-controlled navigation and camera management is especially important in survey areas where topographic maps do not exist, are old, or are of small scale or poor quality. They are 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.

4.7 Selection of sensors for a process study

4.7.1 Data selection criteria

For the selection of the appropriate data, it is necessary to fully understand the information requirements of a specific process study. In a nutshell, the questions to be answered concern coverage and resolution in space, time and spectrum. In addition, cost, availability or acquisition constraints, and quality will also be important. The surface characteristics of the object or objects under study determine which parts of the electromagnetic spectrum will be used for observation (spectral coverage), and whether a few broad bands are needed or many narrow bands (spectral resolution).

The level of detail determines the spatial resolution, whereas the size of the area, or the size of the area of the phenomenon, to be studied determines the spatial coverage, which corresponds to the area covered by one image. Of course, one can use several images to cover the area under study, which is often the case, but mosaicking images increases cost and processing time, and it often causes classification and interpretation problems at the seam between two images. Furthermore, the area of the Earth that can be observed by the sensor to be used is an important spatial aspect. For example, the geostationary MSG-SEVIRI covers only the Western Hemisphere, with very large distortions at the poles.

For the temporal aspect, we have to consider the speed of the process and the duration or the length of the period of observation. The speed of the process determines the frequency of observation within a given time (temporal resolution). When choosing the frequency and time of observation and the moments of observation, however, seasonality should be included in the considerations. For example, glaciers shrink in summer and expand in winter. If one wants to study long-term changes in glaciers, (trend versus cyclic changes), images should be recorded at comparable moments in the year, e.g. end-of-winter, at maximum size, end-of-summer, at minimum size, or images at several moments to get a more accurate estimate of seasonal fluctuations, to be able to separate them from long-term trends.

The temporal coverage needed depends on the duration of the process. For the past, temporal coverage is determined by the image archives of a sensor. Landsat archives date back to 1972, but aerial photographs may be available for many decades back. For the future, i.e. both current and planned satellite missions, continuity in the type of sensor are important. Landsat, NOAA, and Meteosat are examples of series of satellites that are each equipped with similar sensors, which guarantees the continuity of data. Security of continuity of data supply is a major issue for many institutes when deciding on which primary data sources to chose. The JERS-1, with its SAR sensor, has long been a typical example of a "one-off" research mission; JERS-1 operated between 1992 and 1998. It was finally followed up in 2006 with the PALSAR sensor on board ALOS.

The selection of data is further influenced by a number of acquisition constraints. Acquisition of optical data is hampered by clouds, so it is not always possible to acquire an image on a planned date, even if the satellite is in the appropriate orbit. Furthermore, not all sensors can record images continuously, because of power limitations, which means that the number of images recorded per orbit is limited. For stereo air photos, occurrence of optimal Sun elevation angles, resulting in enough shadow for the interpretation of height (but not so much that larger parts of the image are obscured), limits the number of days suitable for image acquisition.

Two quality aspects are especially important for process studies: radiometric quality, and calibration. Because of the shorter dwell time per pixel (ground resolution cell), the radiometric quality of scanners (whiskbroom sensors) is usually less than that of comparable line cameras (pushbroom sensors). Over time, sensors change, so continuous calibration is needed to obtain unbiased observations of the process, and so that trends detected can be attributed to the phenomenon being studied rather than resulting from the aging of the sensor. Furthermore, similar sensors on different platforms in a constellation, or their successors on new platforms, need to be calibrated to make their measurements comparable.

Last but not least, cost plays a major role in image selection for process studies. Although the whole chain of images and processing should be included in cost calculations, in practice the focus tends to be on the cost of the images alone.

To illustrate all these aspects, let us have a look at some typical process studies and the type of data they frequently require. Studies of land processes on a regional or continental scale, for example drought or wildfires, typically use meteorological satellites such as the geostationary Meteosat Second Generation—SEVIRI; the polar orbiting NOAA-AVHRR; or the MODIS sensors of TERRA and AQUA satellites. Spatial resolutions range from 250 m to a few kilometres (depending on location), with the frequency of observations varying from twice a day (MODIS) to every 15 min.

Land use change Studies of land cover change, deforestation and urban expansion use sensors with spatial resolutions between 15 and 60 m, usually with a temporal coverage of more than a decade; observation once or twice a year is usually sufficient. Detailed change studies, focusing on changes within an urban environment or land cover changes in smaller but fragmented areas, use high resolution sensors. Since data from these sensors, with resolutions ranging from less than a metre up to a few metres, are only available for recent years, they are often combined with older aerial photographs.

Monitoring of glaciers Glaciers are dynamic objects: their spatial extent is continuously changing. Monitoring of glaciers can be performed daily, monthly, seasonally or yearly. For studies related to global climate change, one would most likely be interested in data of a longer temporal scale. For example, one could select images of the same season for several years in a row. When selecting images, one should keep weather conditions in mind: dense cloud cover or heavy snow covering the land surface will make delineation of the glacier impossible.

Monitoring of Urban growth The process of urban growth is related to change of land cover type, e.g. construction works, which often take quite some time. Given the typical time scales needed in urban growth studies, annual acquisition of images would be the most appropriate frequency. For studies on urban growth, it does not make sense to acquire images daily.