1

Introduction

1.1 Definition and origins of remote sensing

'Remote sensing' is, broadly but logically speaking, the collection of information about an object without making physical contact with it. This is a simple definition, but too vague to be really useful,[1] so for the purposes of this book we restrict it by confining our attention to the Earth's surface and atmosphere, viewed from above using electromagnetic radiation. This narrower definition excludes such techniques as seismic, geomagnetic and sonar investigations, as well as (for example) medical and planetary imaging, all of which could otherwise reasonably be described as remote sensing, but it does include a broad and reasonably coherent set of techniques, nowadays often described by the alternative name of *Earth Observation*. These techniques, which now have a huge range of applications in the 'civilian' sphere as well as their obvious military uses, make use of information impressed in some way on electromagnetic radiation ranging from ultraviolet to radio frequencies.

The origins of remote sensing can plausibly be traced back to the fourth century BC and Aristotle's *camera obscura* (or, at least, the instrument described by Aristotle in his *Problems*, but perhaps known even earlier). Although significant developments in the theory of optics began to be made in the seventeenth century, and glass lenses were known much earlier than this, the first real advance towards our modern conception of remote sensing came in the first half of the nineteenth century with the invention of photography. For the first time, it became possible to record an image permanently and objectively. Also during the nineteenth century, forms of electromagnetic radiation were discovered beyond the visible part of the spectrum – infrared radiation by Herschel, ultraviolet by Ritter, and radio waves by Hertz – and in 1863

[1] See Campbell (1996, p. 4) for a summary of the main definitions of remote sensing that have been adopted over the last few decades. The term 'remote sensing' itself was first used by the U.S. Office of Naval Research in the 1960s (see Cracknell and Hayes, 1991, p. 1).

Maxwell developed the electromagnetic theory on which so much of our understanding of these phenomena depends.

Aerial photography followed almost immediately on the discovery of the photographic method. The first aerial photograph, unfortunately no longer in existence, was probably made in 1858 by Gaspard Félix Tournachon, taken from a balloon at an altitude of about 80 m. Kites were also soon used, and by 1890 the usefulness of aerial photography was so far recognised that Batut had published a textbook on the subject.

The next step towards what we now recognise as remote sensing was taken with the development of practicable aeroplanes in the early twentieth century. Again, the potential applications were quickly recognised and aerial photographs were recorded from aeroplanes from 1909. Aerial photography was used during the First World War for military reconnaissance, and, during the period between the two World Wars, civilian uses of this technique began to be developed, notably in cartography, geology, agriculture and forestry. Cameras, film and aircraft underwent significant improvements, and stereographic mapping attained an advanced state of development. Also during this period, John Logie Baird, the inventor of television, performed early work on the development of airborne scanning systems capable of transmitting images to the ground. This work was highly confidential, having been carried out on behalf of the French Air Ministry. It was ended by the war and forgotten about until 1985 (Newton, 1989).

The Second World War brought substantial developments to remote sensing. Photographic reconnaissance reached a high state of development – the German invasion of Britain, planned for September 1940, was forestalled by the observation of concentrations of ships along the English Channel. Infrared-sensitive instruments and radar systems were developed. In particular, the Plan Position Indicator used by night bombers was an imaging radar that presented the operator with a 'map' of the terrain, and thus it represented the ancestor of the imaging radar systems discussed in chapter 9.

By the 1950s, false-colour infrared film, originally developed for military use, was finding applications in vegetation mapping, and high-resolution imaging radars were being developed. As these developments continued through the 1960s, sensors began to be placed in space. This was originally part of the programme to observe the Moon, but the advantages of applying the same techniques to observation of the Earth were soon recognised and the first multispectral spaceborne imagery of the Earth was acquired from *Apollo 6*. Although there were earlier unmanned remote sensing satellites,[2] the opening of the modern era of spaceborne remote sensing ought probably to be dated to July 1972 with the successful operation of ERTS, the Earth Resources Technology Satellite, by the U.S. National Aeronautics and Space Administration (NASA). The ERTS was renamed Landsat-1, and the Landsat programme is still continuing – at the time of writing (November 1999), Landsat-7 is producing large quantities of data.

[2] The first was TIROS-1, launched in April 1960.

Since the launch of ERTS in 1972, the number and diversity of spaceborne and airborne remote sensing systems has grown dramatically. A larger range of variables can be measured, and consistent and systematic datasets can be constructed for progressively longer periods of time. The explosive growth in the quantity of data being generated has been matched by growth in the availability of computing resources and the facilities for data storage.

1.2 Applications

The enormous growth in the availability of remotely sensed data over the last four decades has been matched by a fall in the real cost of the data. Nevertheless, it is still clear that use of the data must offer some tangible advantages to justify the cost of acquiring and analysing them. These advantages derive from a number of characteristics of remote sensing. Probably the most important of these is that data can be gathered from a large area of the Earth's surface, or a large volume of the atmosphere, in a short space of time, so that a virtually instantaneous 'snapshot' can be obtained. For example, scanners carried on geostationary meteorological satellites such as METEOSAT can acquire an image of approximately one quarter of the Earth's surface in less than half an hour. When this aspect is combined with the fact that airborne or spaceborne systems can acquire data from locations that would be difficult (slow, expensive, dangerous, politically inconvenient . . .) to measure *in situ*, the potential power of remote sensing becomes apparent. Of course, further advantages derive from the fact that most remote sensing systems generate calibrated digital data that can be fed straight into a computer for analysis.

Remote sensing finds a very wide range of applications, obviously including the area of military reconnaissance in which many of the techniques had their origins. In the non-military sphere, most applications can loosely be categorised as 'environmental', and we can distinguish a range of environmental variables that can be measured. In the atmosphere, these include temperature, precipitation, the distribution and type of clouds, wind velocities, and the concentrations of gases such as water vapour, carbon dioxide, ozone, and so on. Over land surfaces, we can measure tectonic motion, topography, temperature, albedo (reflectance) and soil moisture content, and determine the nature of the land cover in considerable detail, for example by characterising the type of vegetation and its state of health or by mapping man-made features such as roads and towns. Over ocean surfaces, we can measure the temperature, topography (from which the Earth's gravitational field, as well as ocean tides and currents, can be inferred), wind velocity, wave energy spectra, and colour (which is often related to biological productivity by plankton). The 'cryosphere', that part of the Earth's surface covered by snow and ice, can also be studied, giving data on the distribution, condition and dynamical behaviour of snow, sea ice, icebergs, glaciers and ice sheets.

This list of measurable variables, while not complete, is large enough to indicate that there is a correspondingly large number of disciplines to which remote sensing data can be applied. While by no means exhaustive, a list of applications could include the following disciplines: agriculture and crop monitoring, archaeology, bathymetry, cartography, climatology, civil engineering, coastal erosion, disaster monitoring and prediction, forestry, geology, glaciology, oceanography, meteorology, pollution monitoring, snow resources, soil characterisation, urban mapping, water resource mapping and monitoring. It is not really possible to present a detailed cost–benefit analysis in this introduction, partly because, at least until recently, most spaceborne remote sensing operations have been part of national or international space programmes and so their costs have been to some extent hidden.[3] Perhaps it is sufficient to point out that the data available from remote sensing, particularly from spaceborne observations, often cannot be obtained in any other way; that our current understanding of the global climate system is very largely based on spaceborne observations; and that the use of remotely sensed data for disaster warning has already saved many thousands of human lives.

1.3 A systems view of remote sensing

We stated rather briefly in section 1.1 that remote sensing involves the collection of information, carried by electromagnetic radiation, about the Earth's surface or atmosphere. Let us try to expand this statement a little.

First, where does the radiation come from? One major classification of remote sensing systems is into passive systems, which detect naturally occurring radiation, and active systems, which emit radiation and analyse what is sent back to them. Passive systems can be further subdivided into those that detect radiation emitted by the Sun (this radiation consists mostly of ultraviolet, visible light and near-infrared radiation), and those that detect the thermal radiation emitted by all objects that are not at absolute zero (i.e. all objects). For objects at typical terrestrial temperatures, this thermal emission occurs mostly in the infrared part of the spectrum, at wavelengths of the order of 10 μm (the so-called thermal infrared region), although measurable quantities of radiation also occur at longer wavelengths, as far as the microwave part of the spectrum. Active systems can, in principle, use any type of electromagnetic radiation. In practice, however, they are restricted by the transparency of the Earth's atmosphere. This is shown schematically in figure 1.1. The figure shows that there are two main 'windows' in the atmosphere. The first of these includes the visible and infrared parts of the spectrum, between wavelengths of about 0.3 μm and 10 μm, although it does also contain a number of opaque regions; and the second more or less corresponds to the microwave region, between wavelengths of a few millimetres and a few metres. Thus, we can expect that any active system designed to penetrate the

[3] However, a reasonable cost estimate for a typical remote sensing satellite mission might be \$100 million.

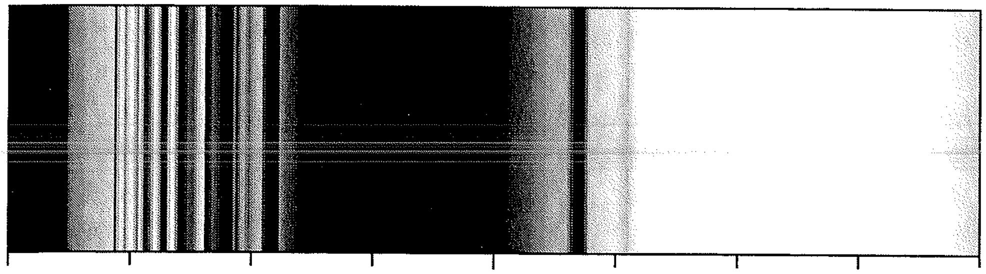

Figure 1.1. Transparency of the Earth's atmosphere as a function of wavelength (schematic). Black regions are opaque, white regions transparent.

Earth's atmosphere will operate in one of these two 'window' regions. Figure 1.2 summarises the main types of remote sensing system on the basis of the classifications we have just outlined.

The sensor, whether it is part of a passive or an active instrument, detects electromagnetic radiation after it has interacted with or been emitted by the 'target' material. In what way can this radiation contain useful information about the target? There are essentially only two variables to describe the radiation that is received: (1) *How much* radiation is detected?[4] (2) *When* does it arrive? The time-structure of the detected radiation is obviously relevant only in the case of active systems where the time-structure of the emitted radiation can be controlled. In this case, it is possible to determine the distance from the sensor to the target, and this is the principle behind various ranging systems such as the laser profiler, lidar, radar altimeter and other types of radar system. In all other cases, the only information we have is the quantity of radiation received at the sensor. If the radiation arises from

	Passive systems			Active systems	
	Reflected sunlight	Thermal emission			
		Infrared	Microwave (radio)	Visible/IR	Microwave (radio)
Non-imaging		Thermal infrared radiometry (6)	Passive microwave radiometry (7)	Laser profiling (8)	Radar altimetry (8) Microwave scatterometry (9)
Imaging	Aerial photography (5) Visible/near-infrared imaging (6)	Thermal infrared imaging (6)	Passive microwave radiometry (7)		Real aperture radar (9) Synthetic aperture radar (9)
Sounding	Ultraviolet backscatter sounding (6)	Thermal infrared sounding (6)	Passive microwave sounding (7)	Lidar (9)	

Figure 1.2. A simple taxonomy of remote sensing systems. The numbers in parentheses refer to the chapters of this book.

[4] The 'quantity' of radiation, however, may need to be qualified with a statement of its polarisation state.

thermal emission, the quantity is characteristic of the temperature of the target material and its emissivity, a property that describes its efficiency at emitting thermal radiation. Otherwise (in the cases of both passive and active systems that measure reflected radiation), the amount of radiation that is received is determined by the amount illuminating the target material, and the target's reflectivity. Thus, we can see that the information about a target material that is directly observable from remote sensing observations is actually rather limited: we can measure its range, its reflectivity, and a combination of its temperature and emissivity. However, these can be measured at different times, over a range of wavelengths and, sometimes, at different polarisation states, and this increase in the diversity of the variables at our disposal is responsible for the large range of indirect observables that was sketched out in section 1.2.

The foregoing discussion has not included the effects of the Earth's atmosphere, except to point out that atmospheric opacity limits the scope for observation of the Earth's surface to the two main atmospheric 'windows'. In fact, as almost any electromagnetic wave propagates through the atmosphere, its characteristics will be somewhat modified. This modification may be troublesome, requiring correction, or advantageous, depending on whether we are more interested in studying the Earth's surface or the atmosphere itself. In general, we can say that if the observation is made at a wavelength at which the atmosphere is opaque, the measured signal will be characteristic of the atmosphere; whereas if the atmosphere is transparent, the data will be characteristic of the surface below.

Once the data have been collected by the sensor, they must be retrieved and analysed. In most, though not all, cases, the data will form an image, by which we mean a two-dimensional representation of the two-dimensional distribution of radiation intensity. Although the familiar photograph can serve as a prototype of an image, the images with which we have to deal in remote sensing are normally digital, so they can conveniently be analysed by computer, and need not be confined to the visible part of the electromagnetic spectrum. For example, an image might represent the radar reflectivity in one or more frequencies or polarisation states, or the thermal emission, as well as the visible or near-infrared reflectivity. Image processing forms an integral part of remote sensing. Typically, this involves several steps. The first is to correct the image so that it has a known geometric correspondence to the Earth's surface and a known calibration, with atmospheric propagation effects removed. At this stage, the image may also be enhanced in various ways, for example by suppressing noise, to increase its intelligibility. The major goal of image processing, however, is the extraction of useful information from the sensor data, based on the brightness values of the image (probably in a number of spectral bands, at a number of different dates, in different polarisation states, etc.) and also on the spatial context. Using the analogy of a colour photograph, we can say that information can be extracted on the basis of colour, texture, shape and spatial context. In the majority of cases, it is necessary or at least desirable to 'train' the process of extracting information from the image using data from known locations. The process can therefore be seen as one of extrapolation from areas that are already known, for example on the basis of field work, to much wider areas. The extrapolation need not be confined to the spatial domain, however, and the analysis of time-series of images for change detection is also an important application of remote sensing.

1.4 Further reading, and how to obtain data

The field of remote sensing is now well served with textbooks, and the interested (or puzzled) reader should be able to find alternative treatments of most of the topics discussed in this book. The 1990s have seen the publication of, amongst other general or introductory texts in remote sensing, *Fundamentals of Remote Sensing and Airphoto Interpretation* by Avery and Berlin (1992), *Introduction to environmental Remote Sensing* by Barrett and Curtis (1992), *Introduction to Remote Sensing* by Campbell (1996), *Introduction to Remote Sensing* by Cracknell and Hayes (1991), *Images of the Earth: a guide to Remote Sensing* by Drury (1998), *Remote Sensing and Image Interpretation* by Lillesand and Kiefer (1994), and *Remote Sensing: principles and interpretations* by Sabins (1996). In addition, useful collections of remote sensing data have been published by, for example, Kramer (1996), Rees (1999) and Ryerson (1998). Amongst older but still very valuable works on remote sensing it is useful to mention *Principles of Remote Sensing* by Curran (1985), *Introduction to the physics and techniques of Remote Sensing* by Elachi (1987), and *Satellite Remote Sensing: an introduction* by Harris (1987).

Scientific journals also represent an important source of information. Articles in the scientific literature are usually aimed at specialists, but the more general reader can often also extract a useful understanding from them, and the journals sometimes also publish review articles. In this book I have provided references to both books and journal articles. The principal English language journals in remote sensing are the *IEEE Transactions on Geoscience and Remote Sensing*, the *International Journal of Remote Sensing*, *Photogrammetric Engineering and Remote Sensing*, and *Remote Sensing of Environment*.

Finally, a few remarks about the Internet may be useful. This can represent a very powerful means of obtaining up-to-date information of all sorts, for example on the operational status of a particular remote sensing satellite, or the latest results from a research group, or access to remote sensing data (indeed, some of the illustrations used in this book have been obtained in this way) or the software needed to process it. As anyone who has grappled with the Internet will know only too well, the problem is usually to locate the information one needs. The well-known 'search engines' can be extremely helpful, as can the collections of 'links' assembled by public-spirited individuals and organisations. I will mention just one source of such links, and that is the web site of the Remote Sensing Society. The interested reader should easily be able to build up a much larger list using this as a starting point. At the time of going to press, its URL (uniform resource locator, i.e. its 'address' on the Internet) was

http://www.the-rss.org