

# Remote Sensing Digital Image Analysis

An Introduction

Bearbeitet von  
John A. Richards

1. Auflage 2012. Buch. xix, 494 S. Hardcover

ISBN 978 3 642 30061 5

Format (B x L): 15,5 x 23,5 cm

Gewicht: 931 g

[Weitere Fachgebiete > EDV, Informatik > Informationsverarbeitung > Bildsignalverarbeitung](#)

Zu [Inhaltsverzeichnis](#)

schnell und portofrei erhältlich bei

  
DIE FACHBUCHHANDLUNG

Die Online-Fachbuchhandlung [beck-shop.de](http://beck-shop.de) ist spezialisiert auf Fachbücher, insbesondere Recht, Steuern und Wirtschaft. Im Sortiment finden Sie alle Medien (Bücher, Zeitschriften, CDs, eBooks, etc.) aller Verlage. Ergänzt wird das Programm durch Services wie Neuerscheinungsdienst oder Zusammenstellungen von Büchern zu Sonderpreisen. Der Shop führt mehr als 8 Millionen Produkte.

## Chapter 2

# Correcting and Registering Images

### 2.1 Introduction

When image data is recorded by sensors on satellites and aircraft it can contain errors in geometry, and in the measured brightness values of the pixels. The latter are referred to as *radiometric errors* and can result from (i) the instrumentation used to record the data (ii) the wavelength dependence of solar radiation and (iii) the effect of the atmosphere.

*Geometric errors* can also arise in several ways. The relative motions of the platform, its scanners and the earth can lead to errors of a skewing nature in an image product. Non-idealities in the sensors themselves, the curvature of the earth and uncontrolled variations in the position, velocity and attitude of the remote sensing platform can all lead to geometric errors of varying degrees of severity.

It is usually important to correct errors in image brightness and geometry. That is certainly the case if the image is to be as representative as possible of the scene being recorded. It is also important if the image is to be interpreted manually. If an image is to be analysed by machine, using the algorithms to be described in [Chaps. 8 and 9](#), it is not always necessary to correct the data beforehand; that depends on the analytical technique being used. Some schools of thought recommend against correction when analysis is based on pattern recognition methods, because correction will not generally improve performance; rather the (minor) discretisation errors that can be introduced into image data by correction procedures may lead to unnecessary interpretation errors. In any case, geometric correction can always be applied to the interpreted product after analysis is complete. Automated interpretation based on library searching or other similarity based methods will always require radiometric correction. Generally, radiometric correction is also required before data fusion operations and when several images of the same region taken at different times are to be compared.

It is the purpose of this chapter to discuss the nature of the radiometric and geometric errors commonly encountered in remote sensing images and to develop computational procedures that can be used for their compensation. The methods to

be presented also find more general application, such as in registering together sets of images of the same region but at different times, and in performing operations such as scale changing and zooming (magnification).

We commence with examining sources of radiometric errors, and methods for their correction, and then move on to problems in image geometry.

## 2.2 Sources of Radiometric Distortion

Mechanisms that affect the measured brightness values of the pixels in an image can lead to two broad types of radiometric distortion. First, the distribution of brightness over an image in a given band can be different from that in the ground scene. Secondly, the relative brightness of a single pixel from band to band can be distorted compared with the spectral reflectance character of the corresponding region on the ground. Both types can result from the presence of the atmosphere as a transmission medium through which radiation must travel from its source to the sensors, and can also be the result of instrumentation effects. The spectral dependence of solar radiation will also affect band to band relativity. We now consider each of these and their correction mechanisms.

## 2.3 Instrumentation Errors

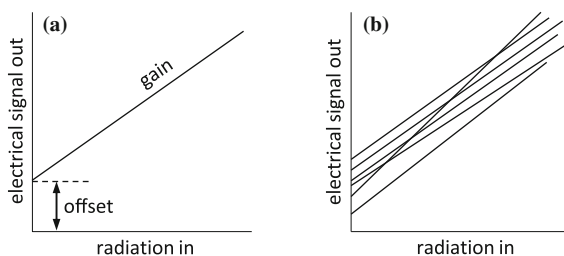
### 2.3.1 Sources of Distortion

Because sets of sensors are used within a band and, obviously, between bands, radiometric errors can arise from calibration differences among the sensors. An ideal radiation detector has a transfer characteristic such as that shown in Fig. 2.1a. It should be linear so that there is a proportionate increase or decrease of signal level with detected radiation. Real detectors will have some degree of non-linearity. There will also be a small signal out, even when there is no radiation in. Historically that is known as *dark current* and is the result of residual electronic noise present in the detector at any temperature other than absolute zero. In remote sensing it is usually called a detector *offset*. The slope of the detector curve is called its *gain*, or sometimes *transfer gain*.

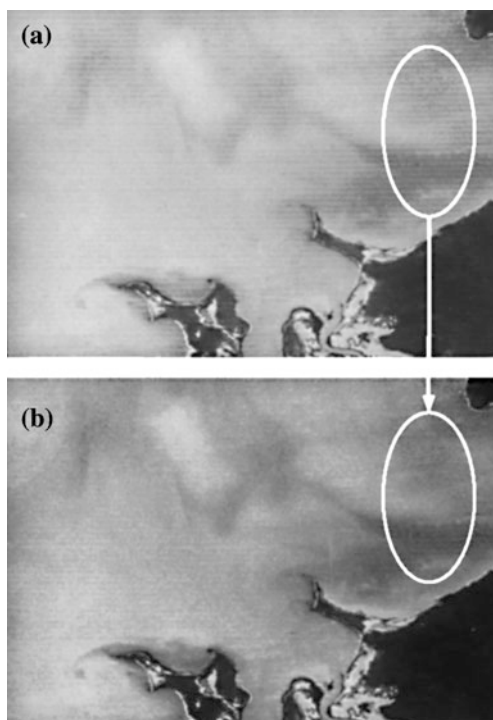
Most imaging devices used in remote sensing are constructed from sets of detectors. In the case of the Landsat ETM+ there are 16 per band. Each will have slightly different transfer characteristics, such as those depicted in Fig. 2.1b. Those imbalances will lead to striping in the across swath direction similar to that shown in Fig. 2.2a.

For push broom scanners, such as the SPOT HRG, there are as many as 12,000 detectors across the swath in the panchromatic mode of operation, so that

**Fig. 2.1** **a** Linear radiation detector transfer characteristic, and **b** hypothetical mismatches in detector characteristics



**Fig. 2.2** Reducing sensor induced striping noise in a Landsat MSS image: **a** original image, and **b** after destriping by matching sensor statistics



longitudinal striping could occur if the detectors were not well matched. For monolithic sensor arrays, such as the charge coupled devices used in the SPOT instruments, that is rarely a problem, compared with the line striping that can occur with mechanical across track scanners that employ discrete detectors.

Another common instrumentation error is the loss of a complete line of data resulting from a momentary sensor or communication link failure, or the loss of signal on individual pixels in a given band owing to instantaneous drop out of a sensor or signal link. Those mechanisms lead to black lines across or along the image, depending on the sensor technology used to acquire the data, or to individual black pixels.

### 2.3.2 Correcting Instrumentation Errors

Errors in relative brightness, such as the within-band line striping referred to above and as shown in Fig. 2.2a for a portion of a Landsat Multispectral Scanner (MSS) image, can be rectified to a great extent in the following way. First, it is assumed that the detectors used for data acquisition in each band produce signals statistically similar to each other. In other words, if the means and standard deviations are computed for the signals recorded by each of the detectors over the full scene then they should almost be the same. This requires the assumption that statistical detail within a band doesn't change significantly over a distance equivalent to that of one scan covered by the set of the detectors (474 m for the six scan lines of Landsats 1, 2, 3 MSS for example). For most scenes this is usually a reasonable assumption in terms of the means and standard deviations of pixel brightness, so that differences in those statistics among the detectors can be attributed to the gain and offset mismatches illustrated in Fig. 2.1b

Sensor mismatches of this type can be corrected by calculating pixel mean brightness and standard deviation using lines of image data known to come from a single detector. In the case of Landsat MSS that will require the data on every sixth line to be used. In a like manner five other measurements of mean brightness and standard deviation are computed for the other five MSS detectors. Correction of radiometric mismatches among the detectors can then be carried out by adopting one sensor as a standard and adjusting the brightnesses of all pixels recorded by each other detector so that their mean brightnesses and standard deviations match those of the standard detector. That operation, which is commonly referred to as *destriping*, can be implemented by the operation

$$y = \frac{\sigma_d}{\sigma_i} x + m_d - \frac{\sigma_d}{\sigma_i} m_i \quad (2.1)$$

where  $x$  is the original brightness for a pixel and  $y$  is its new (destriped) value in the band being corrected;  $m_d$  and  $\sigma_d$  are the reference values of mean brightness and standard deviation, usually those of a chosen band, and  $m_i$  and  $\sigma_i$  are the signal mean and standard deviation for the detector under consideration. Sometimes an independent reference mean and standard deviation is used. That allows a degree of contrast enhancement to be imposed during the destriping operation.

Figure 2.2 shows the result of applying (2.1) to the signals of the remaining five detectors of a Landsat Multispectral Scanner (MSS) image, after having chosen one as a reference. As seen, the result is good but not perfect, partly because the signals are being matched only on the basis of first and second order statistics. A better approach is to match the detector histograms using the methodology of Sect. 4.5.<sup>1</sup> It is also possible to correct errors in an observed image by using optimisation to

---

<sup>1</sup> This approach is demonstrated in M.P. Weinreb, R. Xie, I.H. Lienesch and D.S. Crosby, Destriping GOES images by matching empirical distribution functions, *Remote Sensing of Environment*, vol. 29, 1989, pp. 185–195, and M. Wegener, Destriping multiple sensor imagery by improved histogram matching, *Int. J. Remote Sensing*, vol. 11, no. 5, May 1990, pp. 859–875.

match it to an assumed error-free image model,<sup>2</sup> and to use sub-space methods when the dimensionality is high.<sup>3</sup> More complex methods, however, are generally less suitable with large numbers of detectors.

Correcting lost lines of data or lost pixels can be carried out by averaging over the neighbouring pixels—using the lines on either side for line drop outs or the set of surrounding pixels for pixel drop outs. This is called *infilling* or sometimes *in-painting*.

## 2.4 Effect of the Solar Radiation Curve and the Atmosphere on Radiometry

We now examine the effect of environmental conditions on the radiometric character of recorded image data. To help focus on the important aspects, consider a hypothetical surface which will reflect all of the incident sunlight at all wavelengths. Assume, further, that there is no atmosphere above the surface, as depicted in Fig. 2.3a. A detector capable of taking many spectral samples will record the solar spectrum as shown.<sup>4</sup>

Now suppose there is a normal terrestrial atmosphere in the path between the sun, the surface and the detector. The spectrum recorded will be modified by the extent to which the atmosphere selectively absorbs the radiation. There are well-known absorption features caused mainly by the presence of oxygen, carbon dioxide and water vapour in the atmosphere, and they appear in the recorded data as shown in Fig. 2.3b. The atmosphere also scatters the solar radiation, further complicating the signal received at the sensor. This reduces the solar energy that strikes the surface and travels to the sensor; energy also scatters from the atmosphere itself to the sensor superimposing onto the desired signal. We consider those additional complications in Sect. 2.6.

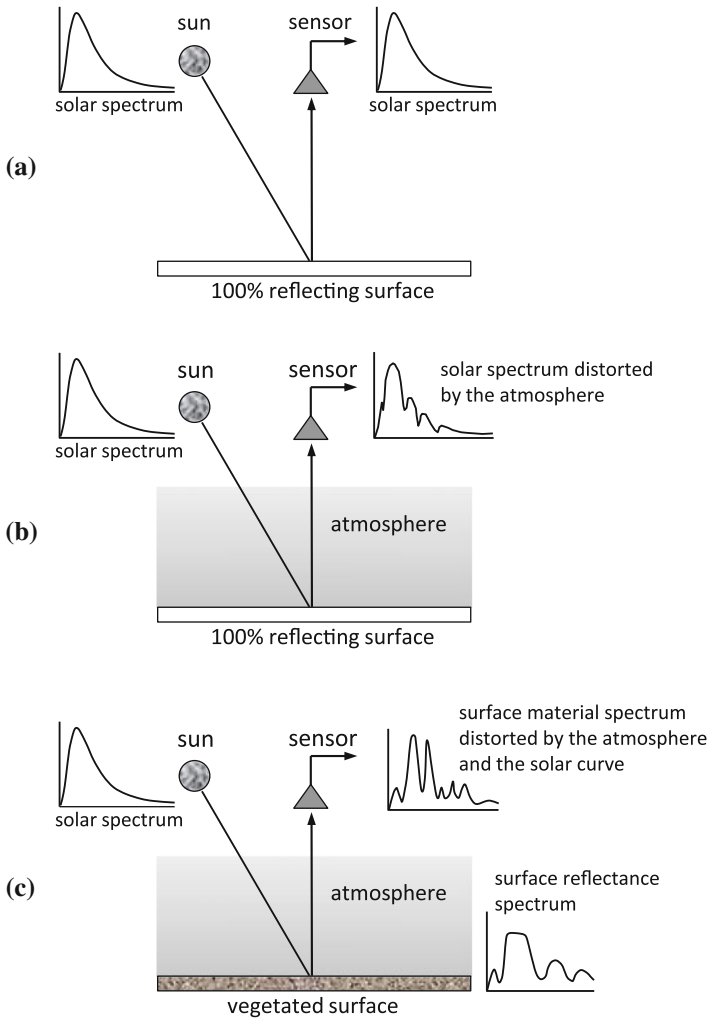
Figure 2.3c shows how the reflectance spectrum of a *real* surface might appear. The spectrum recorded is a combination of the actual spectrum of the real surface, modulated by the influence of the solar curve and distorted by the atmosphere. In order to be able to recover the true radiometric character of the image we need to correct for those effects.

---

<sup>2</sup> See H. Shen and L. Zhang, A MAP-based algorithm for destriping and inpainting of remotely sensed images, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 5, May 2009, pp. 1492–1502, and M. Bouali and S. Ladjal, Towards optimal destriping of MODIS data using a unidirectional variance model, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 8, August 2011, pp. 2924–2935.

<sup>3</sup> See N. Acito, M. Diani and G. Corsini, Subspace-based striping noise reduction in hyperspectral images, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 4, April 2011, pp. 1325–1342.

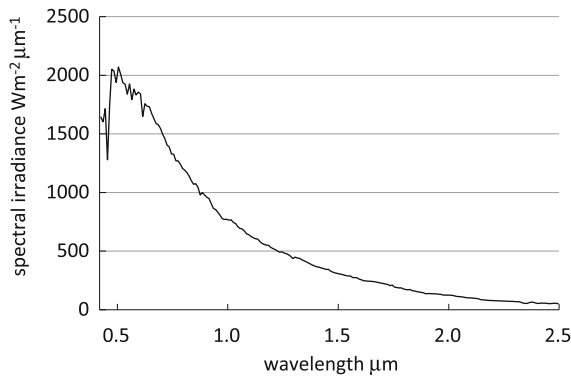
<sup>4</sup> If the spectral resolution of the detector were sufficiently fine then the recorded solar spectrum would include the Fraunhofer absorption lines associated with the gases in the solar atmosphere: See P.N. Slater, *Remote Sensing: Optics and Optical Systems*, Addison Wesley, Reading Mass., 1980.



**Fig. 2.3** Distortion of the surface material reflectance spectrum by the spectral dependence of the solar curve and the effect of the atmosphere: **a** detection of the solar curve from a perfectly reflecting surface in the absence of an atmosphere, **b** effect of the atmosphere on detecting the solar curve, **c** detection of the real spectrum distorted by the atmosphere and the solar curve

## 2.5 Compensating for the Solar Radiation Curve

In a very simple way the wavelength dependence of the radiation falling on the earth's surface can be compensated by assuming that the sun is an ideal black body and able to be described by the behaviour of the Planck radiation law shown in Fig. 1.3. For broad spectral resolution sensors that is an acceptable approach. For images recorded by instrumentation with fine spectral resolution it is important to account for departures



**Fig. 2.4** Measured solar spectral irradiance of the sun above the earth's atmosphere over the wavelength range common in optical remote sensing; plotted, at lower spectral resolution, from the data in F.X. Kneizys, E.P. Shettle, L.W. Abreu, J.H. Chetwynd, G.P. Anderson, W.O. Gallery, J.E.A. Selby and S.A. Clough, *Users Guide to LOWTRAN7*, AFGL-TR-0177, Environmental Research Paper No 1010, 1988, which can be found at [www.star.nesdis.noaa.gov/smcd/spb/calibration/avhrr/solrad.html](http://www.star.nesdis.noaa.gov/smcd/spb/calibration/avhrr/solrad.html)

from black body behaviour, effectively modelling the real emissivity of the sun, and using that to normalise the recorded image data. Most radiometric correction procedures compensate for the solar curve using the actual wavelength dependence measured above the atmosphere, such as that shown in Fig. 2.4.

## 2.6 Influence of the Atmosphere

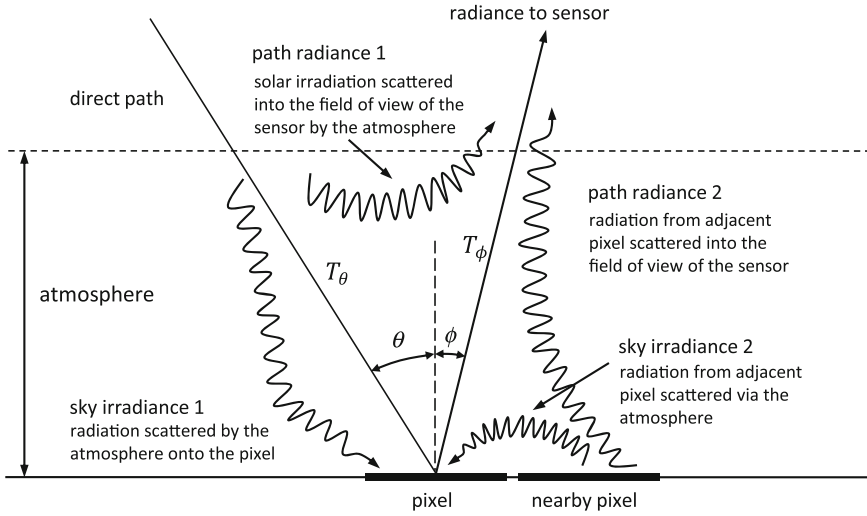
We now examine how solar irradiation produces the measured signal from a single pixel, using the mechanisms identified in Fig. 2.5. It is important, first, to define radiometric quantities in order to allow the correction equations to be properly formulated.

The sun is a source of energy that emits at a given rate of joules per second, or watts. That energy radiates through space isotropically in an inverse square law fashion so that at a particular distance the sun's emission can be measured as watts per square metre ( $\text{Wm}^{-2}$ ), given as the power emitted divided by the surface area of a sphere at that distance. This power density is called *irradiance*, a property that can be used to describe the strength of any emitter of electromagnetic energy.

The power density scattered from the earth in a particular direction is defined by density per solid angle. This quantity is called *radiance* and has units of watts per square metre per steradian ( $\text{Wm}^{-2}\text{sr}^{-1}$ ). If the surface is perfectly diffuse then the incident solar irradiance is scattered uniformly into the upper hemisphere, i.e. equal amounts are scattered into equal cones of solid angle.

The emission of energy by bodies such as the sun is wavelength dependent, as seen in Fig. 1.3, so that the term *spectral irradiance* can be used to describe how much power density is available in incremental wavebands across the wavelength range;





**Fig. 2.5** Effect of the atmosphere on solar radiation illuminating a pixel and reaching a sensor

that is the quantity plotted in Fig. 2.4. Spectral irradiance is measured in  $\text{Wm}^{-2}\mu\text{m}^{-1}$ . Similarly, *spectral radiance* is measured in  $\text{Wm}^{-2}\mu\text{m}^{-1}\text{sr}^{-1}$ .

Suppose in the absence of the atmosphere the solar spectral irradiance at the earth is  $E_\lambda$ . If the solar zenith angle (measured from the normal to the surface) is as shown in Fig. 2.5 then the spectral irradiance (spectral power density) at the earth's surface is  $E_\lambda \cos\theta$ . This gives an available irradiance between wavelengths  $\lambda_1$  and  $\lambda_2$  of

$$E_{os} = \int_{\lambda_1}^{\lambda_2} E_\lambda \cos\theta d\lambda \text{ Wm}^{-2}$$

For most instruments the wavebands used are sufficiently narrow that we can assume

$$E_{os} = E_{\Delta\lambda} \cos\theta \Delta\lambda = E(\lambda) \cos\theta \text{ Wm}^{-2} \quad (2.2)$$

in which  $\Delta\lambda = \lambda_2 - \lambda_1$  and  $E_{\Delta\lambda}$  is the average spectral irradiance over that bandwidth, centred on the wavelength  $\lambda = (\lambda_2 + \lambda_1)/2$ .  $E(\lambda) = E_{\Delta\lambda} \Delta\lambda$  is the solar irradiance above the atmosphere at wavelength  $\lambda$ .

Suppose the surface has a reflectance  $R$  in that narrow band of wavelengths, which describes the proportion of the incident irradiance that is scattered. If the surface is diffuse then the total radiance  $L$  scattered into the upper hemisphere, and available for measurement, is

$$L = E(\lambda) \cos\theta R / \pi \text{ Wm}^{-2}\text{sr}^{-1} \quad (2.3)$$

in which the divisor  $\pi$  accounts for the upper hemisphere of solid angle. This equation relates to the ideal case of no atmosphere.

When an atmosphere is present there are two effects which must be taken into account that modify (2.3). They are the scattering and absorption by the particles in the atmosphere, for which compensation is needed when correcting imagery. Absorption by atmospheric molecules is a selective process that converts incoming energy into heat; molecules of oxygen, carbon dioxide, ozone and water attenuate the radiation very strongly in certain wavebands.

There are two broad scattering mechanisms. The first is scattering by the air molecules themselves. That is called Rayleigh scattering, which depends on the inverse fourth power of the wavelength. The other is called aerosol or Mie scattering and is the result of scattering of radiation from larger particles such as those associated with smoke, haze and fumes. Those particulates are of the order of one tenth to ten wavelengths. Mie scattering is also wavelength dependent, although not as strongly as Rayleigh scattering; it is approximately inversely proportional to wavelength. When the atmospheric particulates become much larger than a wavelength, such as those common in fogs, clouds and dust, the wavelength dependence disappears.

In a clear ideal atmosphere Rayleigh scattering is the only mechanism present. It accounts for the blueness of the sky. Because the shorter (blue) wavelengths are scattered more than the longer (red) wavelengths we are more likely to see blue when looking in any direction in the sky. Likewise the reddish appearance of sunset is also caused by Rayleigh scattering. That is the result of the long atmospheric path the radiation has to follow at sunset during which most short wavelength radiation is scattered away from direct line of sight, relative to the longer wavelengths.

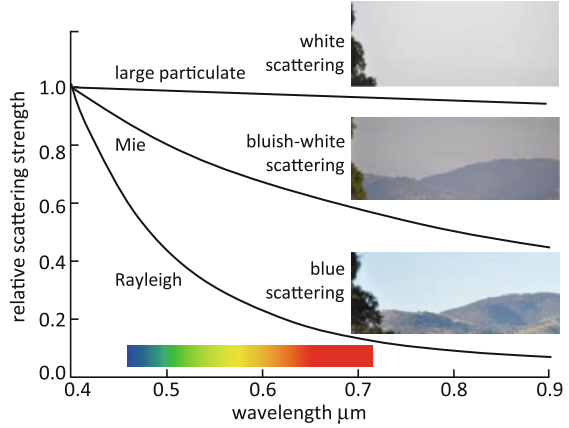
Fogs and clouds appear white or bluish-white owing to the (almost) non-selective scattering caused by the larger particles. Figure 2.6 shows the typical scattering characteristics of different atmospheres.

We are now in the position to include the effect of the atmosphere on the radiation that ultimately reaches a sensor. We will do this by reference to the mechanisms shown in Fig. 2.5, commencing with the incoming solar radiation. They are identified by name:

*Transmittance.* In the absence of an atmosphere transmission of the available solar irradiance to the surface at any wavelength is 100%. However, because of scattering and absorption, not all of the solar radiation reaches the ground. The amount that does, relative to that for no atmosphere, is called the transmittance. Let this be denoted  $T_\theta$  in which the subscript indicates its dependence on the zenith angle of the source which determines the path length through the atmosphere. In a similar way there is an atmospheric transmittance  $T_\phi$ , between the point of reflection and the sensor. In general both are wavelength dependent.

*Sky irradiance.* Because the radiation is scattered on its travel down through the atmosphere a particular pixel will be irradiated by energy on the direct path in Fig. 2.5 and by energy scattered from atmospheric constituents. The path for the latter is undefined and diffuse, and is referred to as sky irradiance component 1. A pixel can also receive energy that has been reflected from surrounding pixels and then scattered downwards by the atmosphere. That is

**Fig. 2.6** Atmospheric and particulate scattering



the sky irradiance component 2 shown in Fig. 2.5. We call the total sky irradiance at the pixel  $E_D$ .

*Path radiance.* Again because of scattering, radiation can reach the sensor from adjacent pixels and also via diffuse scattering of the incoming radiation that is actually scattered towards the sensor by atmospheric constituents before it reaches the ground. Those two components constitute path radiance, which is denoted as  $L_p$ .

Having defined these mechanisms we are now in the position to determine how the radiance measured by the sensor is affected by the presence of the atmosphere. First, the total irradiance at the earth's surface now becomes, instead of (2.2),

$$E_G = E(\lambda)T_\theta \cos\theta + E_D \text{ Wm}^{-2}$$

where, for simplicity, it has been assumed that the diffuse sky irradiance  $E_D$  is not a function of wavelength in the waveband of interest. The radiance resulting from this total irradiance of the pixel is thus

$$L = \{E(\lambda)T_\theta \cos\theta + E_D\}R/\pi \text{ Wm}^{-2}\text{sr}^{-1}$$

Emerging from the atmosphere the total radiance detected by the sensor is composed of that term, reduced by atmospheric transmittance on the upward path, plus the path radiance  $L_p$ , to give

$$L = T_\phi \{E(\lambda)T_\theta \cos\theta + E_D\}R/\pi + L_p \text{ Wm}^{-2}\text{sr}^{-1} \quad (2.4)$$

This equation gives the relationship between the radiance measured at the sensor  $L$  and the reflectance of the surface material  $R$  in a given waveband, assuming all the other quantities can be modelled or measured. Sometimes the path radiance term is written as  $L_p = E(\lambda)R_A/\pi$  in which  $R_A$  is called the reflectance of the atmosphere.<sup>51</sup>

<sup>5</sup> See *ACORN 4.0 Users Guide, Stand Alone Version*, Analytical Imaging and Geophysics LLC, Boulder, Colorado, 2002.

If the diffuse sky irradiance term  $E_D$  is neglected (2.4) becomes

$$L(\lambda) = T(\lambda)R(\lambda)E(\lambda)\cos\theta/\pi + L_p(\lambda) \text{ Wm}^{-2}\text{sr}^{-1} \quad (2.5)$$

where the explicit dependence on wavelength (the band or channel centre) is now shown for all quantities;  $E(\lambda)$  is the available solar irradiance in the channel of interest,  $T(\lambda)$  is the transmittance of the total atmospheric path,  $L(\lambda)$  is the radiance at the detector and  $L_p(\lambda)$  is the path radiance. Equation (2.5) is used in many atmospheric correction procedures (see Sect. 2.9).

Image data products are expressed in digital numbers on a scale set by the radiometric resolution of the sensor<sup>6</sup>; 8 bit data is in the range 0–255, 10 bit data is in the range 0–1023 and 12 bit data is in the range 0–4095. The relationship between the detected radiance and the corresponding digital number ( $DN$ ) in the image product can be expressed

$$L = \kappa DN + L_{min} \text{ Wm}^{-2}\text{sr}^{-1} \quad (2.6a)$$

in which the sensor gain term  $\kappa$  is

$$\kappa = (L_{max} - L_{min})/DN_{max} \text{ Wm}^{-2}\text{sr}^{-1} \text{ per digital value} \quad (2.6b)$$

$DN_{max}$  is the highest possible digital count for the sensor. Values for  $L_{max}$ ,  $L_{min}$  and  $DN_{max}$  in each waveband are usually available from the sensor operator, allowing the digital data to be expressed in radiance. This is a necessary step before the correction of atmospheric errors.

## 2.7 Effect of the Atmosphere on Remote Sensing Imagery

The result of atmospheric distortion of the signal recorded by a sensor depends, to an extent, on the spectral resolution of the instrument. We consider broad waveband systems first, such as Landsat ETM+ and SPOT HRG.

One effect of scattering is that fine detail in image data will be obscured. Consequently, in applications where one depends on the limit of sensor resolution, it is important to take steps to correct for atmospheric effects.

It is important also to consider the effects of the atmosphere on systems with wide fields of view in which there will be an appreciable difference in atmospheric path length between nadir and the extremities of the swath. That will be significant with satellite missions such as NOAA.

Finally, and perhaps most importantly, because both Rayleigh and Mie scattering are wavelength dependent, the effects of the atmosphere will be different in the different wavebands of a given sensor system. In the case of the Landsat Thematic Mapper the visible blue band (0.45–0.52  $\mu\text{m}$ ) can be affected

---

<sup>6</sup> See Appendix B.

appreciably by comparison to the middle infrared band (1.55–1.75  $\mu\text{m}$ ). That leads to a loss in calibration over the set of brightnesses for a particular pixel.

In high spectral resolution systems, such as hyperspectral sensors, the effect of the atmosphere is complicated by the presence of the absorption lines superimposed by water vapour and other atmospheric constituents. We examine suitable correction mechanisms in [Sect. 2.9](#).

## 2.8 Correcting Atmospheric Effects in Broad Waveband Systems

Correcting images from broad waveband sensors (typically multispectral), to remove as much as possible the degrading effects of the atmosphere, entails modelling the scattering and gross absorption processes to establish how they determine the transmittances of the signal paths, and the components of sky irradiance and path radiance. When those quantities are available they can be used in [\(2.4\)](#) or [\(2.5\)](#), and [\(2.6\)](#) to relate the digital numbers  $DN$  for the pixels in each band of data to the true reflectance  $R$  of the surface being imaged. An instructive example of how this can be done is given by Forster<sup>7</sup> for the case of Landsat MSS data; he also gives source material and tables to assist in the computations.

Forster considers the case of a Landsat 2 MSS image in the wavelength range 0.8 to 1.1  $\mu\text{m}$  (near infrared; then called band 7) acquired at Sydney, Australia on 14th December, 1980 at 9:05 am local time. At the time of overpass the atmospheric conditions were

temperature	29 °C
relative humidity	24% measured at 30 m above sea level
atmospheric pressure	1004 mbar
visibility	65 km

Based on the equivalent mass of water vapour in the atmosphere (computed from temperature and humidity measurements) the absorbing effect of water molecules was computed. That is the only molecular absorption mechanism considered significant over the broad waveband involved. The measured value for visibility was used to estimate the effect of Mie scattering. That was combined with the known effect of Rayleigh scattering at that wavelength to give the so-called total normal optical thickness of the atmosphere. Its value for this example is  $\tau = 0.15$ . The transmittance of the atmosphere at an angle of incidence  $\theta$  is given by

$$T = \exp(-\tau \sec \theta)$$

---

<sup>7</sup> B.C. Forster, Derivation of atmospheric correction procedures for Landsat MSS with particular reference to urban data. *Int. J. Remote Sensing*, vol. 5, no. 5, 1984, pp. 799–817.

For a solar zenith angle of  $38^\circ$  (at overpass) and a nadir viewing satellite we find

$$T_\theta = 0.827$$

$$T_\phi = 0.861$$

In the waveband of interest Forster shows that the solar irradiance at the earth's surface in the absence of an atmosphere is  $E_0 = 256 \text{ Wm}^{-2}$ . He further computes the total global irradiance at the earth's surface as  $186.6 \text{ Wm}^{-2}$ . Noting in (2.4) that the term in brackets is the global irradiance, and using the relevant values of  $T_\theta$  and  $\cos\theta$ , this gives the total diffuse sky irradiance as  $19.6 \text{ Wm}^{-2}$ —i.e. about 10% of the global irradiance for this example.

Using correction algorithms given by Turner and Spencer,<sup>8</sup> which account for Rayleigh and Mie scattering and atmospheric absorption, Forster computes the path radiance for this example as

$$L_p = 0.62 \text{ Wm}^{-2}\text{sr}^{-1}$$

so that (2.4) becomes for band 7

$$L_7 = 0.274 \times 186.6 R_7 + 0.62 \text{ Wm}^{-2}\text{sr}^{-1}$$

$$\text{i.e.} \quad L_7 = 51.5 R_7 + 0.62 \text{ Wm}^{-2}\text{sr}^{-1} \quad (2.7)$$

At the time of overpass it was established that for the band 7 sensor on Landsat  $2 L_{\max} = 39.1 \text{ Wm}^{-2}\text{sr}^{-1}$  and  $L_{\min} = 1.1 \text{ Wm}^{-2}\text{sr}^{-1}$ , while  $DN_{\max} = 63$  (6 bit data) so that, from (2.6b)

$$\kappa = 0.603 \text{ Wm}^{-2}\text{sr}^{-1} \text{ per digital value}$$

From (2.6a) we thus have

$$L_7 = 0.603 DN_7 + 1.1 \text{ Wm}^{-2}\text{sr}^{-1}$$

which, when combined with (2.7), gives the corrected reflectance in band 7 as

$$R_7 = 0.0118 DN_7 + 0.0094$$

$$\text{or, as a percentage,} \quad R_7 = 1.18 DN_7 + 0.94\%$$

Similar calculations for the visible red band (band 5) give

$$R_5 = 0.44 DN_5 + 0.5\%$$

---

<sup>8</sup> R.E. Turner and M.M. Spencer, Atmospheric model for the correction of spacecraft data, *Proc. 8th Int. Symposium on Remote Sensing of the Environment*, Ann Arbor, Michigan, 1972, pp. 895–934.

## 2.9 Correcting Atmospheric Effects in Narrow Waveband Systems

Correcting image data from narrow waveband sensors (hyperspectral) requires careful modelling of the differential absorption characteristics of atmospheric constituents and their scattering effects, and compensating for the solar curve as discussed in Sect. 2.5. The high spectral resolution means that fine atmospheric absorption features will be detected and may be confused with those of the ground cover type being imaged if not removed.

Correction in these cases consists of the following steps:

1. Conversion of raw recorded *DN* values to *radiance*, unless the product supplied is already in radiance form.
2. Compensating for the shape of the solar spectrum as outlined in Sect. 2.5. The measured radiances are divided by solar irradiances above the atmosphere to obtain the *apparent* reflectances of the surface.
3. Compensating for atmospheric gaseous transmittances, and molecular and aerosol scattering. Simulating these atmospheric effects allows the *apparent* reflectances to be converted to *scaled* surface reflectances.
4. Converting surface reflectances to *real* surface reflectances after considering any topographic effects. If topographic data is not available, real reflectance is taken to be identical to scaled reflectance under the assumption that the surfaces of interest are Lambertian; that is the assumption made by many correction procedures and is used here.

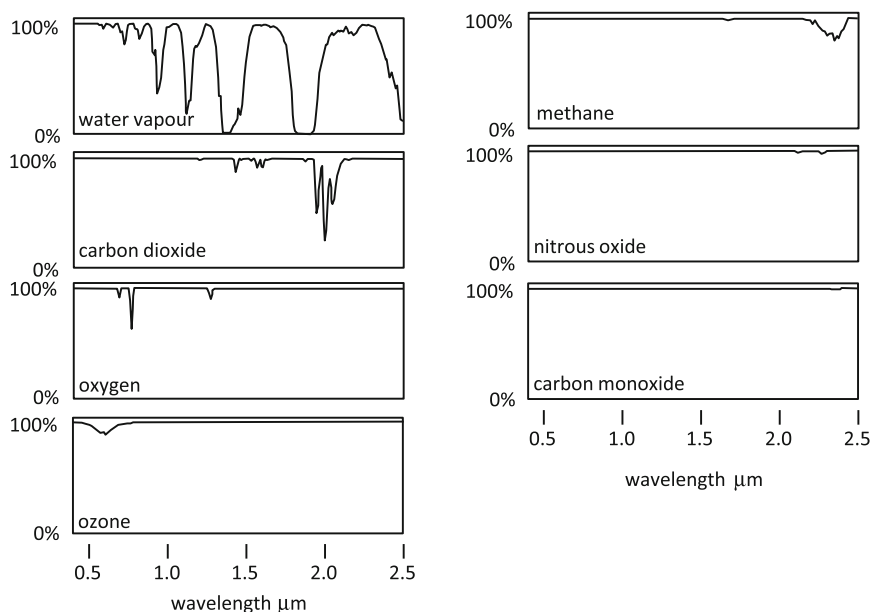
We need now to consider how the third step can be performed. To do so we use the simplified expression of (2.5), which requires information on the absorptive and scattering properties of significant atmospheric constituents. Absorption enters via the transmittance  $T(\lambda)$  and scattering via both the transmittance  $T(\lambda)$  and the path radiance term  $L_p$ .

Over the years several data bases listing specific absorption characteristics of atmospheric gaseous components have been compiled. The most extensive, and that which is used by several atmospheric correction models in remote sensing, is HITRAN. Although its heritage can be traced back to 1973, successive refinements have led to an extensive compilation of the effects of a great number of important and less significant atmospheric constituents<sup>9</sup>; detailed information on the use of HITRAN and its development is available separately.<sup>10</sup>

Not all of the molecular constituents covered in HITRAN are significant when correcting high spectral resolution remote sensing imagery. The most important in the range relevant to optical remote sensing, 0.4–2.5  $\mu\text{m}$ , are  $\text{H}_2\text{O}$ ,  $\text{CO}_2$ ,  $\text{O}_3$ ,  $\text{N}_2\text{O}$ ,

<sup>9</sup> L.S. Rothman and 42 others, The HITRAN 2008 molecular spectroscopic database, *J. Quantitative Spectroscopy and Radiative Transfer*, vol. 110, 2009, pp. 533–572.

<sup>10</sup> See [www.cfa.harvard.edu/HITRAN/](http://www.cfa.harvard.edu/HITRAN/).



**Fig. 2.7** Indicative transmission spectra of the seven most significant atmospheric constituents, adapted from Figs. 2 and 3 of B. Gao, K.B. Heidebrecht and A.F.H. Goetz, *loc cit*; used with permission of Elsevier; the water vapour curve is for a tropical atmosphere

CO, CH<sub>4</sub> and O<sub>2</sub>.<sup>11</sup> Their transmission characteristics (referred to as transmission spectra) are illustrated in Fig. 2.7.

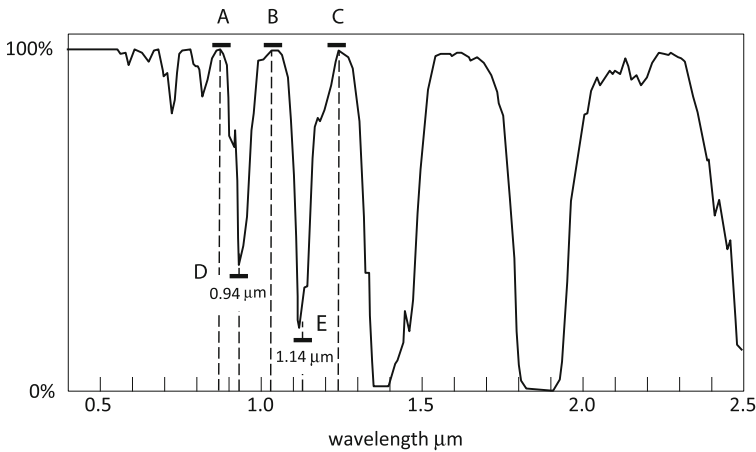
Apart from ozone, which varies with latitude and season, but which can be modelled as a constant effect for a given image, all of CO<sub>2</sub>, N<sub>2</sub>O, CO, CH<sub>4</sub>, O<sub>2</sub> can be considered relatively constant from image to image, and their absorption characteristics modelled<sup>12</sup> and used to correct for their absorbing effects on hyperspectral imagery.

Correction for the effects of water vapour is more complex because water in the atmosphere changes with humidity, and can vary across a scene. Ideally it would be good to estimate the water vapour in the atmospheric path for each individual pixel so that each pixel can have its reflectivity corrected for water vapour absorption and scattering separately. Fortunately, with fine spectral resolution systems, that turns out to be possible through examining the resonant water absorption dips evident in Fig. 2.7, and reproduced in Fig. 2.8 with further relevant information added. The depths of the minima at wavelengths of 0.94 and 1.14  $\mu\text{m}$  depend on the amount of water vapour in the atmospheric path for the

<sup>11</sup> B.C. Gao, K.B. Heidebrecht and A.F.H. Goetz, Derivation of scaled surface reflectance from AVIRIS data, *Remote Sensing of Environment*, vol. 44, 1993, pp. 165–178.

<sup>12</sup> Many correction methodologies use the narrow band transmittance model in W. Malkmus, Random Lorentz band model with exponential-tailed S line intensity distribution function, *J. Optical Society of America*, vol. 57, 1967, pp. 323–329.





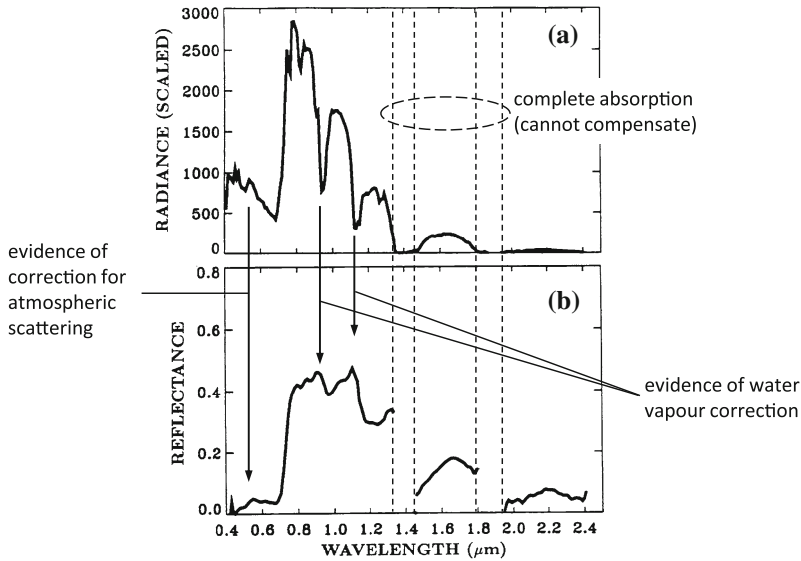
**Fig. 2.8** Using absorption features in the water spectrum to estimate atmospheric water content; the average signal at D is divided by the averages over the ranges A and B, and the average signal at E is divided by the averages of the ranges B and C to give two measurements for estimating water content

relevant pixel. We can assess the quantity of water in the column by comparing the depths of either of those minima (as averages of a set of bands around those wavelengths) with the 100% transmission level shown; 100% transmission occurs in the water vapour windows near 0.865, 1.025 and 1.23  $\mu\text{m}$ , so bands, or averages over sets of bands, near those wavelengths can be used to provide the 100% reference levels. Once the depth of a water absorption minimum has been estimated, usually by taking the ratio of the radiance at the minimum to the average radiance of the 100% transmission bands either side, a model is used to generate the water content in the path from the sun to the sensor, via the respective pixel. That allows the corresponding transmission coefficient to be derived and the path radiance contributed by the atmospheric water content to be determined.

Several packages are available that implement radiometric correction based on the processes just described. One of the earliest was ATREM (Atmosphere Removal) developed at the University of Colorado.<sup>13</sup> It accounts for the seven atmospheric constituents noted above, using the ratio technique of Fig. 2.8 to account and correct for atmospheric water vapour. Atmospheric scattering is incorporated using the 5S and 6S radiative transfer codes.<sup>14</sup> The vegetation spectrum for a pixel that has been atmospherically corrected using ATREM is shown in Fig. 2.9.

<sup>13</sup> *Atmosphere Removal Program (ATREM), Version 3.1 Users Guide*, Centre for the Study of Earth from Space, University of Colorado, 1999.

<sup>14</sup> D. Tanre, C. Deroo, P. Duhaut, M. Herman, J.J. Morcrette, J. Perbos and P.Y. Deschamps, *Simulation of the Satellite Signal in the Solar Spectrum (5S) Users Guide*, Laboratoire d'Optique Atmosphérique, Université S.T. de Lille, 1986, and <http://6s.ltdri.org/index.html> (6S users site).



**Fig. 2.9** Correction of the raw spectrum of a vegetation pixel in which key features are evident; underlying diagram reprinted, with permission from Elsevier, from B.C. Gao, K. B. Heidebrecht and A.F.H. Goetz, Derivation of scaled surface reflectance from AVIRIS data, *Remote Sensing of Environment*, vol. 44, 1993, pp. 165–178

ATREM is no longer available. Current correction programs tend to be based on, and are refinements of, MODTRAN4.<sup>15</sup> MODTRAN4, unlike ATREM allows the sky irradiance term 2 and path radiance term 2 in Fig. 2.5 to be incorporated into the atmospheric correction process. Two recent approaches that build on MODTRAN4 are ACORN<sup>16</sup> (Atmospheric Correction Now) and FLAASH<sup>17</sup> (Fast Line-of-Sight Atmospheric Analysis of Hyperspectral Cubes). A comparison of these packages will be found in Kruse,<sup>18</sup> while Gao et al.<sup>19</sup> review developments in correction algorithms and indicate where improvements are required.

<sup>15</sup> A. Berk, G.P. Anderson, L.S. Bernstein, P.K. Acharya, H. Dothe, M.W. Matthew, S.M. Adler-Golden, J.H. Chetwynd, Jr., S.C. Richtsmeier, B. Pukall, C.L. Allred, L.S. Jeong, and M.L. Hoke, MODTRAN4 Radiative Transfer Modeling for Atmospheric Correction, *Proc. SPIE Optical Stereoscopic Techniques and Instrumentation for Atmospheric and Space Research III*, vol. 3756, July 1999.

<sup>16</sup> ACORN 4.0 Users Guide, Stand Alone Version, loc. cit.

<sup>17</sup> S.M. Alder-Golden, M.W. Matthew, L.S. Bernstein, R.Y. Levine, A. Berk, S.C. Richtsmeier, P.K. Acharya, G.P. Anderson, G. Felde, J. Gardner, M. Hike, L.S. Jeong, B. Pukall, J. Mello, A. Ratkowski and H.H. Burke, Atmospheric correction for short wave spectral imagery based on MODTRAN4, *Proc. SPIE Imaging Spectrometry*, vol. 3753, 1999, pp. 61–69.

<sup>18</sup> F.A. Kruse, Comparison of ATREM, ACORN and FLAASH atmospheric corrections using low-altitude AVIRIS data of Boulder, CO, *Proc. 13th JPL Airborne Geoscience Workshop*, Pasadena, CA, 2004.

<sup>19</sup> B.C. Gao, M.J. Montes, C.O. Davis and A.F.H. Goetz, Atmospheric correction algorithms for hyperspectral remote sensing data of land and oceans, *Remote Sensing of Environment*, Supplement 1, Imaging Spectroscopy Special Issue, vol. 113, 2009, pp. S17–S24.

## 2.10 Empirical, Data Driven Methods for Atmospheric Correction

Several approximate techniques are available for atmospheric correction that depend directly on the measurements of the recorded image data. These are important when detailed data on the atmosphere, and particularly the water vapour content, are not available. The most common are considered here.

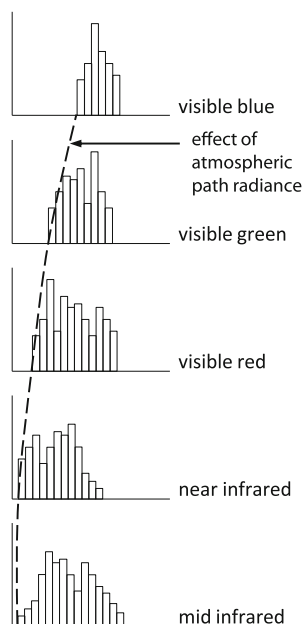
### 2.10.1 Haze Removal by Dark Subtraction

Frequently, detailed correction for the scattering and absorbing effects of the atmosphere is not required in broad waveband systems. Neither can detailed correction be implemented when the necessary ancillary information, such as visibility and relative humidity, is not readily available. In those cases, if the effect of the atmosphere is judged to be a problem, approximate correction can be carried out in the following manner. Effectively, it just corrects for the path radiance term  $L_p$  in (2.4); some commercial image processing software systems use this method to account for path radiance before other procedures are applied to compensate for atmospheric absorption effects.

It makes the assumption that each band of data for a given scene should contain some pixels at or close to zero brightness value but that atmospheric effects, and especially path radiance, has added a constant level to each pixel in each band. Consequently, if histograms are taken of the bands (graphs of the number of pixels as a function of brightness value) the lowest significant occupied brightness value will be non-zero as shown in Fig. 2.10. Also, because path radiance varies as  $\lambda^{-\alpha}$  (with  $\alpha$  between 0 and 4 depending upon the extent of Mie scattering) the lowest occupied brightness value will be further from the origin for the shorter wavelengths, as depicted. Approximate correction amounts, first, to identifying the amount by which each histogram is apparently shifted in brightness from the origin and then subtracting that amount from each pixel brightness in that band.

It is clear that the effect of atmospheric scattering as implied in the histograms of Fig. 2.10 is to lift the overall brightness value of an image in each band. In the case of a colour composite product (see Sect. 3.2.1) this will appear as a whitish-bluish haze. Following correction in the manner just described—often called *dark subtraction*—the haze will be removed and the dynamic range of image intensity will be improved. Consequently this approach is also frequently referred to as *haze removal*.

**Fig. 2.10** Illustrating the effect of path radiance resulting from atmospheric scattering



### 2.10.2 The Flat Field Method

The Flat Field method<sup>20</sup> depends on locating a large, spectrally uniform area in an image, such as sand or clouds (a “spectrally” flat field) and computing its average radiance spectrum. It is assumed that the recorded shape and absorption features present in the spectrum are caused by solar and atmospheric effects since, in their absence, the spectrum should be flat. The reflectance of each image pixel is then corrected by dividing the spectrum of the pixel by the average radiance spectrum of the flat field.

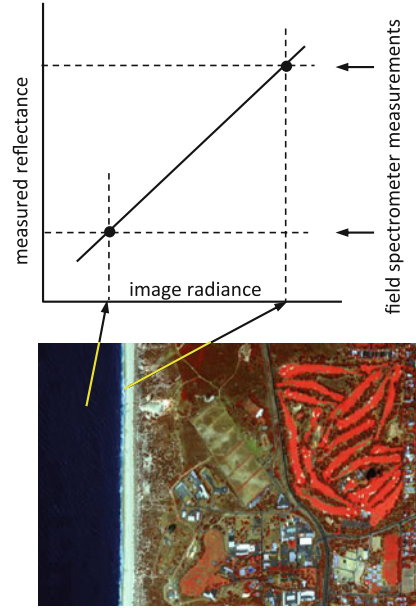
### 2.10.3 The Empirical Line Method

In this approach<sup>21</sup> two spectrally uniform targets in the image, one dark and one bright, are identified; their actual reflectances are then determined by field or laboratory measurements. The radiance spectra for each target are extracted from

<sup>20</sup> D.A. Roberts, Y. Yamaguchi and R.J.P. Lyon, Comparison of various techniques for calibration of AIS data, *Proc. 2nd AIS Workshop*, JPL Publication 86–35, Jet Propulsion Laboratory, Pasadena CA, 1986.

<sup>21</sup> D.A. Roberts, Y. Yamaguchi and R.J.P. Lyon, Calibration of Airborne Imaging Spectrometer data to percent reflectance using field spectral measurements, *Proc. 19th Int. Symposium on Remote Sensing of Environment*, Ann Arbor, Michigan, 21–25 October 1985.

**Fig. 2.11** Illustrating the Empirical Line Method



the image and then mapped to the actual reflectances using linear regression techniques. The gain and offset so derived for each band are then applied to *all* pixels in the image to calculate their reflectances, as illustrated in Fig. 2.11. While this is an appealing technique and the computational load is manageable, it does require field or laboratory reflectance data to be available.

#### 2.10.4 Log Residuals

This method is based on an assumed linear relationship between radiance and reflectance of the form<sup>22</sup>

$$x_{i,n} = I_n S_i R_{i,n} \quad (2.8)$$

where  $x_{i,n}$  ( $i = 1 \dots K, n = 1 \dots N$ ) is the radiance for pixel  $i$  in waveband  $n$  and  $R_{i,n}$  is the reflectance to be found.  $S_i$  accounts for the effect of topography, different for each pixel but assumed to be constant for all wavelengths.  $I_n$  accounts for wavelength dependent illumination, including the solar curve and atmospheric transmittance, which is assumed to be independent of pixel.  $K$  and  $N$  are the total number of the pixels in the image and the total number of bands, respectively.

<sup>22</sup> A.A. Green and M.D. Craig, Analysis of Airborne Imaging Spectrometer data with logarithmic residuals, *Proc. 1st AIS Workshop*, JPL Publication 85-41, Jet Propulsion Laboratory, Pasadena CA, 8-10 April 1985, pp. 111-119.

For reasons which will become clear shortly divide the measured radiance  $x_{i,n}$  by its geometric mean over the wavebands and its geometric mean over the pixels. Here we denote the geometric mean of a quantity  $x$  with respect to the index  $n$  by  $\mathcal{G}_n(x)$  so that we have

$$z_{i,n} = \frac{x_{i,n}}{\mathcal{G}_n(x_{i,n})\mathcal{G}_i(x_{i,n})} \quad (2.9)$$

Substituting from (2.8), this becomes

$$z_{i,n} = \frac{I_n S_i R_{i,n}}{\mathcal{G}_n(I_n) S_i \mathcal{G}_n(R_{i,n}) I_n \mathcal{G}_i(S_i) \mathcal{G}_i(R_{i,n})}$$

$$\text{i.e.} \quad z_{i,n} = \frac{R_{i,n}}{\mathcal{G}_n(I_n) \mathcal{G}_i(S_i) \mathcal{G}_n(R_{i,n}) \mathcal{G}_i(R_{i,n})} \quad (2.10)$$

Now  $\mathcal{G}_n(I_n)$  is independent of pixel and thus is a constant in (2.10); likewise  $\mathcal{G}_i(S_i)$  is independent of band and is also a constant. Therefore, to within a multiplicative constant,  $z_{i,n}$  defined on the basis of measured radiance in (2.9) is an expression involving surface reflectance which is of the same form as that involving measured radiance, and is independent of both the illumination conditions and the effect of topography. Accordingly, if we used (2.9) then the result can be considered to be a scaled reflectance.

We now take the logarithm of (2.9), to give

$$\log z_{i,n} = \log x_{i,n} - \log \mathcal{G}_n(x_{i,n}) - \log \mathcal{G}_i(x_{i,n})$$

$$\text{i.e.} \quad \log z_{i,n} = \log x_{i,n} - \frac{1}{N} \sum_{n=1}^N \log x_{i,n} - \frac{1}{K} \sum_{i=1}^K \log x_{i,n} \quad (2.11)$$

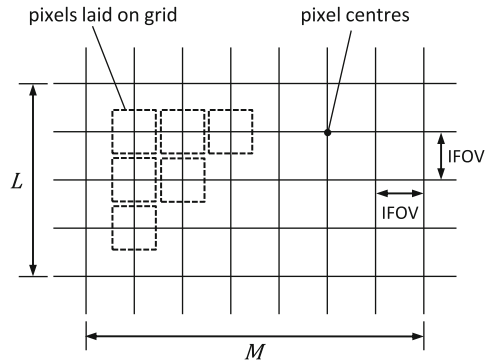
This is the expression used for the log residuals method. It produces the logarithm of an expression equivalent to scaled reflectance, and is thus independent of topographic and illumination effects.

## 2.11 Sources of Geometric Distortion

There are potentially many more sources of geometric distortion of images than radiometric distortion, and their effects can be quite severe. Some are more important with aircraft platforms whereas others are a greater problem for satellite based sensors. They can be related to a number of factors, including

- the rotation of the earth during image acquisition
- variations in platform altitude, attitude and velocity
- the wide field of view of some sensors

**Fig. 2.12** The display grid used to build up an image from the digital data stream of pixels generated by a sensor



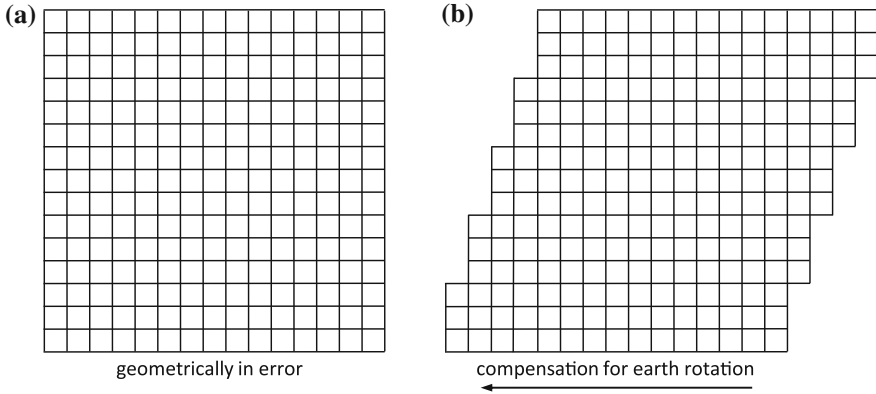
- the curvature of the earth
- the finite scan rate of some sensors
- sensor non-idealities

In the following sections we discuss the nature of the distortions that arise from these effects and means by which they can be compensated.

To appreciate why geometric distortion occurs and its manifestation in imagery, it is important to envisage how an image is formed from sequential lines of image data. If one imagines that a particular sensor records  $L$  lines of  $M$  pixels each then it would be natural to form the image by laying the  $L$  lines down successively one under the other. If the IFOV of the sensor has an aspect ratio of unity—i.e. the pixels are the same size along and across the scan—then this is the same as arranging the pixels for display on a square grid, such as that shown in Fig. 2.12. The grid intersections are the pixel positions and the spacing between the grid points is equal to the sensor's IFOV.

## 2.12 The Effect of Earth Rotation

Sensors that record one line of data at a time across the image swath will incur distortion in the recorded image product as a result of the rotation of the earth during the finite time required to record a full scene. During the frame (or scene) acquisition time the earth rotates from west to east so that a pixel imaged at the end of the frame would have been further to the west when recording started. Therefore if the lines of pixels recorded were arranged for display in the manner of Fig. 2.12 the later lines would be erroneously displaced to the east in terms of the terrain they represent. To give the pixels their correct positions relative to the ground it is necessary to offset the bottom of the image to the west by the amount by which the ground has moved during image acquisition, with all intervening lines displaced proportionately as depicted in Fig. 2.13. The amount by which the image has to be skewed to the west at the end of the frame depends on the relative velocities of the satellite and earth, and the length of the image frame recorded.



**Fig. 2.13** Effect of earth rotation on image geometry when data is acquired as scan lines: **a** image constructed according to Fig. 2.12 in which the pixels are arranged on a square grid, **b** offset of successive groups of lines to the west to correct for earth rotation during image acquisition

An example is presented here for Landsat 7. The angular velocity of the satellite is  $\omega_o = 1.059 \text{ mrad s}^{-1}$  so that a nominal  $L = 185 \text{ km}$  frame on the ground is scanned in

$$t_s = \frac{L}{\omega_o r_e} = 27.4 \text{ s}$$

where  $r_e$  is the radius of the earth (6.37816 Mm). The surface velocity of the earth is

$$v_e = \omega_e r_e \cos \varphi$$

in which  $\varphi$  is latitude and  $\omega_e$  is the earth rotational velocity of  $72.72 \text{ } \mu\text{rad s}^{-1}$ . At Sydney, Australia  $\varphi = 33.8^\circ$  so that

$$v_e = 385.4 \text{ ms}^{-1}$$

During the frame acquisition time the surface of the earth at Sydney moves to the east by

$$\Delta x_e = v_e t_s = 10.55 \text{ km}$$

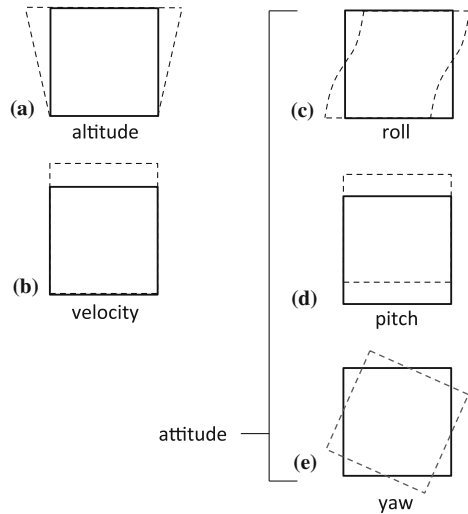
This is 6% of the 185 km frame, which is quite severe and certainly noticeable. Since the satellite does not pass directly north–south, this figure has to be adjusted by the path inclination angle. At Sydney for Landsat 7 this is approximately  $11^\circ$  so that the effective sideways movement of the earth is actually

$$\Delta x = \Delta x_e \cos 11^\circ = 10.34 \text{ km}$$

If steps are not taken to correct for the effect of earth rotation during Landsat 7 image acquisition the image will exhibit about 6% skew distortion to the east.



**Fig. 2.14** Effect of platform position and attitude variations on the region of the earth being imaged, for variations that are slow compared with image acquisition



### 2.13 The Effect of Variations in Platform Altitude, Attitude and Velocity

Variations in the elevation or altitude of a remote sensing platform lead to a scale change at constant angular IFOV and field of view; the effect is illustrated in Fig. 2.14a for an increase in altitude with travel at a rate that is slow compared with a frame acquisition time. Similarly, if the platform forward velocity changes, a scale change occurs in the along-track direction. That is depicted in Fig. 2.14b again for a change that occurs slowly. For a satellite platform orbit velocity variations can result from orbit eccentricity and the non-sphericity of the earth.

Platform attitude changes can be resolved into yaw, pitch and roll during forward travel. These lead to image rotation, along track and across track displacement as noted in Fig. 2.14c–e. The effects in the recorded imagery can be understood by again referring to Fig. 2.12 while looking at the diagrams in Fig. 2.14. For example, while Fig. 2.14a shows that the field of view of a sensor broadens with rising platform height, mapping the recorded pixels onto the grid of Fig. 2.12 will lead to an apparent compression of detail compared with that at lower altitudes.

Attitude variations in aircraft remote sensing systems can be quite significant owing to the effects of atmospheric turbulence. Those variations can occur over a short time, leading to localised distortions in aircraft scanner imagery.<sup>23</sup> Aircraft roll can be partially compensated in the data stream. That is made possible by having a data window that defines the swath width; the window is made smaller

<sup>23</sup> For an extreme example see Fig. 3.1 in G. Camps-Valls and L. Bruzzone, eds., *Kernel Methods for Remote Sensing Data Analysis*, John Wiley & Sons, Chichester UK, 2009.

than the complete scan of data over the sensor field of view. A gyroscope mounted on the sensor is then used to move the position of the data window along the total scan line as the aircraft rolls. Pitch and yaw are generally not corrected unless the sensor is mounted on a three axis stabilized platform.

While these variations can be described mathematically, at least in principle, a knowledge of the platform ephemeris is needed for their magnitudes to be computed.

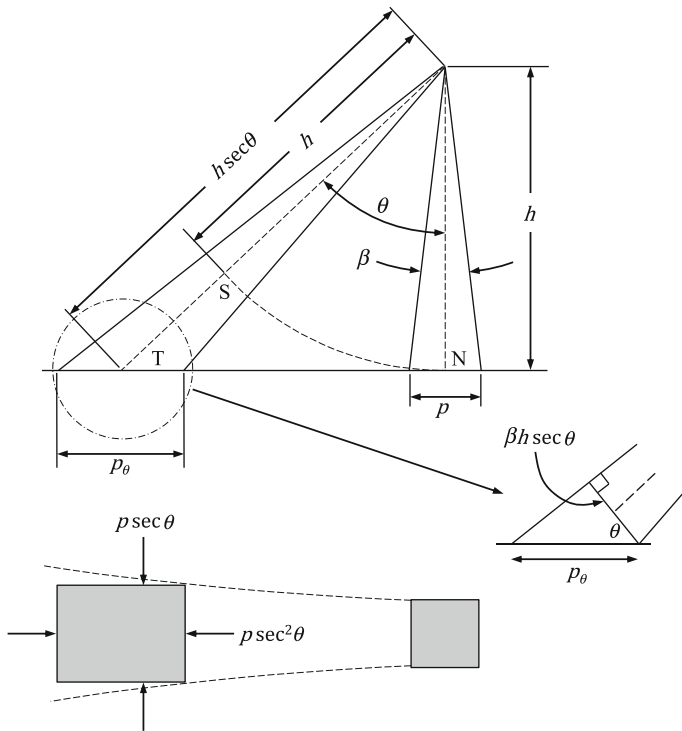
## 2.14 The Effect of Sensor Field of View: Panoramic Distortion

For scanners used on spacecraft and aircraft remote sensing platforms the angular IFOV is constant. As a result the effective pixel size on the ground is larger at the extremities of the scan than at nadir, as illustrated in Fig. 2.15. If the IFOV is  $\beta$  and the pixel dimension at nadir is  $p$  then its dimension in the scan direction at a scan angle  $\theta$  as shown is

$$p_\theta = \beta h \sec^2 \theta = p \sec^2 \theta \quad (2.12)$$

where  $h$  is altitude. Its dimension across the scan line is  $p \sec \theta$ . For small values of  $\theta$  distortion in pixel size is negligible. For Landsat 7 the largest value of  $\theta$  is approximately  $7.5^\circ$  so that  $p_\theta = 1.02p$ . The effect can be quite severe for systems with larger fields of view, such as MODIS and aircraft scanners. For an aircraft scanner with  $\text{FOV} = 80^\circ$  the distortion in pixel size along the scan line is  $p_\theta = 1.70p$ , i.e. the region on the ground measured at the extremities of the scan is 70% larger laterally than the region sensed at nadir. When the image data is arranged to form an image, as in Fig. 2.12, the pixels are all written as the same pixel size on a display device. While the displayed pixels are equal across the scan line the equivalent ground areas covered are not. This gives a compression of the image data towards its edges.

There is a second, related distortion introduced with wide field of view systems concerning pixel positions across the scan line. The scanner records pixels at constant angular increments and these are displayed on a grid of uniform centres, as in Fig. 2.12. However the spacings of the effective pixels on the ground increase with scan angle. For example, if the pixels are recorded at an angular separation equal to the IFOV of the sensor then at nadir the pixels centres are spaced  $p$  apart. At a scan angle  $\theta$  the pixel centres will be spaced  $p \sec^2 \theta$  apart as can be found from Fig. 2.15. By placing the pixels on a uniform display grid the image will suffer an across track compression. Again the effect for small angular field of view systems will be negligible in terms of the relative spacing of adjacent pixels. However when the effect is aggregated to determine the location of a pixel at the swath edge relative to nadir the error can be significant. This can be determined by computing the arc SN in Fig. 2.15, S being the position to which the pixel at T would appear to be moved if the data is arrayed uniformly. It can be shown readily that  $\text{SN}/\text{TN} = \theta/\tan \theta$  that being the degree of across track scale distortion. In the case of Landsat 7  $(\theta/\tan \theta)_{\max} = 0.9936$ . This indicates that

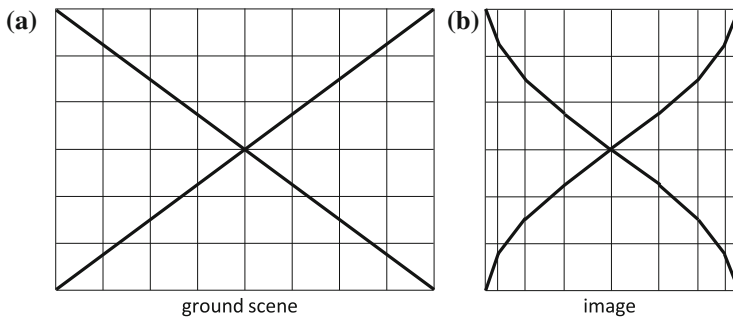


**Fig. 2.15** Effect of scan angle on pixel size at constant angular instantaneous field of view

a pixel at the swath edge (92.5 km from the sub-nadir point) will be 314 m out of position along the scan line compared with the ground, if the pixel at nadir is in its correct location.

These panoramic effects lead to an interesting distortion in the geometry of large field of view systems. To see this consider the uniform mesh shown in Fig. 2.16a which represents a region being imaged. The cells in the grid could be considered to be large fields on the ground. Because of the compression caused by displaying equal-sized pixels on a uniform grid as discussed above, the uniform mesh will appear as shown in Fig. 2.16b. Recall that image pixels are recorded with a constant IFOV and at a constant angular sampling rate. The number of pixels recorded therefore over the outer grid cells in the scan direction will be smaller than over those near nadir. In the along-track direction there is no variation of pixel spacing or density with scan angle as this is established by the forward motion of the platform, although pixels near the swath edges will contain information in common owing to the overlapping IFOV.

Linear features such as roads at an angle to the scan direction as shown in Fig. 2.16 will appear bent in the displayed image data because of the across scan compression effect. Owing to the change in shape, the distortion is frequently



**Fig. 2.16** Along scan line compression incurred in constant IFOV and constant scan rate sensors, leading to S bend distortion

referred to as S-bend distortion and can be a common problem with aircraft line scanners. Clearly, not only linear features are affected; rather the all of the detail near the swath edges is distorted.

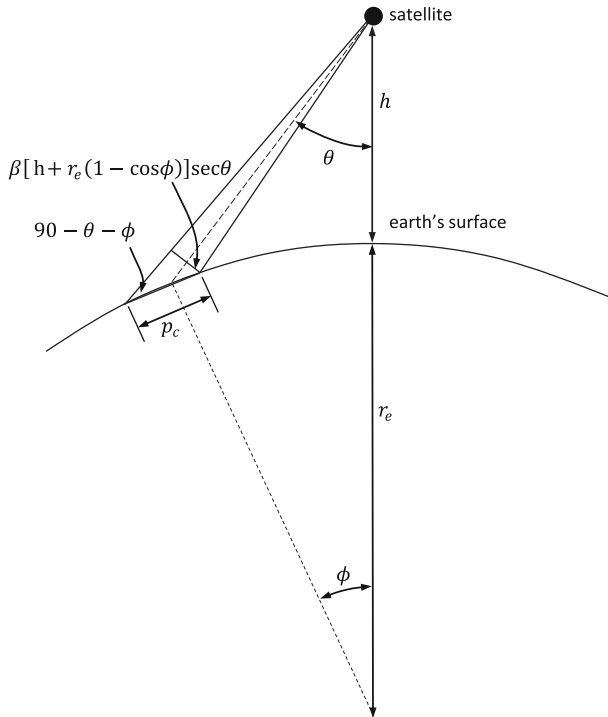
## 2.15 The Effect of Earth Curvature

Aircraft scanning systems, because of their low altitude and thus small absolute swath width, are usually not affected by earth curvature. Neither are small FOV space systems such as Landsat and SPOT, again because of the narrowness of their swaths. However wide swath width spaceborne imaging systems are affected. For MODIS, with a swath width of 2330 km and an altitude of 705 km, it can be shown that the deviation of the earth's surface from a plane amounts to less than 1% over the swath, which seems insignificant. However it is the inclination of the earth's surface over the swath that causes the greater effect. At the edges of the swath the area of the earth's surface viewed at a given angular IFOV is larger than if the curvature of the earth were ignored; that exaggerates the panoramic effect treated in the previous section. The increase in pixel size can be computed by reference to the geometry of Fig. 2.17. The pixel dimension in the across track direction normal to the direction of the sensor is  $\beta[h + r_e(1 - \cos \phi)] \sec \theta$  as shown. The effective pixel size on the inclined earth's surface is then

$$p_c = \beta[h + r_e(1 - \cos \phi)] \sec \theta \sec(\theta + \phi) \quad (2.13)$$

where  $\beta h$  is the pixel size at nadir and  $\phi$  is the angle subtended at the centre of the earth. Note that this expression reduces to (2.12) as  $\phi \rightarrow 0$ , i.e. if earth curvature is negligible.

Using the NOAA satellite as an example  $\theta = 54^\circ$  at the edge of the swath and  $\phi = 12^\circ$ . Equation (2.12) shows that the effective pixel size in the along scan



**Fig. 2.17** Effect of earth curvature on the size of a pixel in the across track direction

direction is 2.89 times larger than that at nadir when earth curvature is ignored but, from (2.13), is 4.94 times that at nadir when the effect of earth curvature is included. Thus earth curvature introduces a significant additional compressive distortion in the image data acquired by satellites such as NOAA when an image is constructed on the uniform grid of Fig. 2.12. The effect of earth curvature in the along-track direction is negligible.

## 2.16 Geometric Distortion Caused by Instrumentation Characteristics

Depending on its style of operation, the sensor used for image acquisition can also introduce geometric distortion into the recorded image data. Here we look at three typical sources of distortion encountered with instruments that build up an image by scanning across the flight line, such as with Landsat, AVHRR, MODIS and some aircraft scanners.

### ***2.16.1 Sensor Scan Nonlinearities***

Line scanners that make use of rotating mirrors, such as the MODIS and AVHRR, have a scan rate across the swath that is constant, to the extent that the scan motor speed is constant. Systems that use an oscillating mirror, such as the Landsat Thematic Mapper, incur some nonlinearity near the swath edges in their scan angle versus time characteristic owing to the need for the mirror to slow down and change direction. That will lead to a displacement distortion of the recorded pixel data in the along-track direction.<sup>24</sup>

### ***2.16.2 Finite Scan Time Distortion***

Mechanical line scanners require a finite time to scan across the swath. During this time the satellite is moving forward, skewing the recorded image data in the along-track direction. As an illustration of the magnitude of the effect, the time required to record one Landsat MSS scan line of data is 33 ms. In this time the satellite travels forward by 213 m at its equivalent ground velocity of  $6.461 \text{ km s}^{-1}$ . The end of the scan line is advanced by that amount compared with its start. The Landsat Thematic Mapper compensates for this error source by using a scan skew corrector mirror.

### ***2.16.3 Aspect Ratio Distortion***

The aspect ratio of an image—its scale vertically compared with its scale horizontally—can be distorted by mechanisms that lead to overlapping IFOVs. The most notable example of this in the past occurred with the Landsat Multispectral Scanner. By design, samples were taken across a scan line too quickly compared with the IFOV. That led to pixels having 56 m centres even though sampled with an IFOV of 79 m. Consequently, the effective pixel size in MSS imagery is  $79 \text{ m} \times 56 \text{ m}$  rather than square. As a result, if the pixels recorded by the Multispectral Scanner are displayed on the square grid of Fig. 2.12 the image will be too wide for its height when related to the corresponding region on the ground. The magnitude of the distortion is  $79/56 = 1.411$  which is quite severe and must be corrected for most applications. Similar distortion can occur with aircraft scanners if the velocity of the aircraft is not matched to the scanning rate of the sensor. Either under-scanning or over-scanning can occur distorting the along-track scale of the image.

---

<sup>24</sup> For Landsat multispectral scanner products this can lead to a maximum displacement in pixel position compared with a perfectly linear scan of about 395 m; see P. Anuta, Geometric correction of ERTS-1 digital MSS data, *Information Note 103073*, Laboratory for Applications of Remote Sensing, Purdue University West Lafayette, Indiana, 1973.

## 2.17 Correction of Geometric Distortion

There are two techniques that can be used to correct the various types of geometric distortion present in digital image data. One is to model the nature and magnitude of the sources of distortion and use the model to establish correction formulas. That approach is effective when the types of distortion are well characterised, such as that caused by earth rotation. The second method depends on establishing mathematical relationships between the addresses of pixels in an image and the corresponding coordinates of those points on the ground (via a map).<sup>25</sup> Those relationships can be used to correct image geometry irrespective of the analyst's knowledge of the source and type of distortion. This approach will be treated first since it is the most commonly used and, as a technique, is independent of the platform used for data acquisition. Correction by mathematical modelling is discussed later. Note that each band of image data has to be corrected. Since it can usually be assumed that the bands are well registered to each other, steps taken to correct one band can be used on all remaining bands.

## 2.18 Use of Mapping Functions for Image Correction

An assumption made in this procedure is that there is available a map of the region covered by the image, which is correct geometrically. We then define two Cartesian coordinate systems as shown in Fig. 2.18. One describes the location of points in the map  $(x, y)$  and the other defines the location of pixels in the image  $(u, v)$ . Suppose that the two coordinate systems can be related via a pair of mapping functions, such that

$$u = f(x, y) \quad (2.14a)$$

$$v = g(x, y) \quad (2.14b)$$

If these functions were known then we could locate a point in the image knowing its position on the map. In principle, the reverse is also true. With this knowledge we could build up a geometrically correct version of the image in the following manner. First we define a grid over the map to act as the grid of pixel centres for the corrected image. This grid is parallel to, or could be, the map coordinate grid described by latitudes and longitudes, UTM coordinates, and so on. For simplicity we will refer to that grid as the *display grid*; by definition it is geometrically correct. We then move over the display grid pixel centre by pixel centre and use

---

<sup>25</sup> For a recent, comprehensive treatment of image correction and registration see J. Le Moigne, N. S. Netanyahu and R. D. Eastman, eds., *Image Registration for Remote Sensing*, Cambridge University Press, Cambridge, 2011.

the mapping functions above to find the pixel in the image corresponding to each display grid position. Those pixels are then placed on the display grid. At the conclusion of the process we have a geometrically correct image built up on the display grid using the original image as a source of pixels.

While the process is straightforward there are some practical difficulties that we must address. First, we do not know the explicit form of the mapping functions in (2.14). Secondly, even if we did, for a given display grid location they may not point exactly to a pixel in the image. In such a case some form of interpolation will be required.

### 2.18.1 Mapping Polynomials and the Use of Ground Control Points

Since explicit forms for the mapping functions in (2.14) are not known they are usually approximated by polynomials of first, second or third degree. In the case of second degree (or order)

$$u = a_0 + a_1x + a_2y + a_3xy + a_4x^2 + a_5y^2 \quad (2.15a)$$

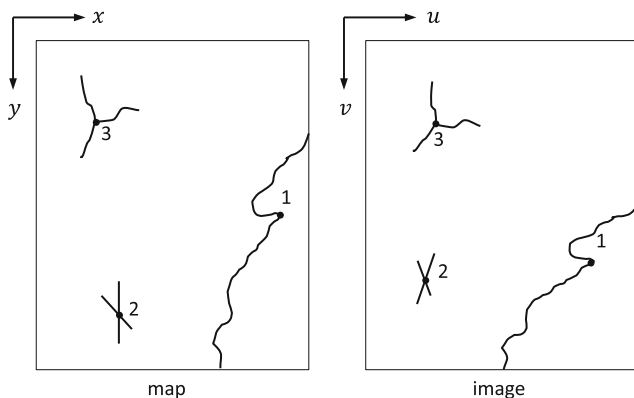
$$v = b_0 + b_1x + b_2y + b_3xy + b_4x^2 + b_5y^2 \quad (2.15b)$$

Sometimes orders higher than three are used but care must be taken to avoid the introduction of errors worse than those to be corrected. That will be discussed later.

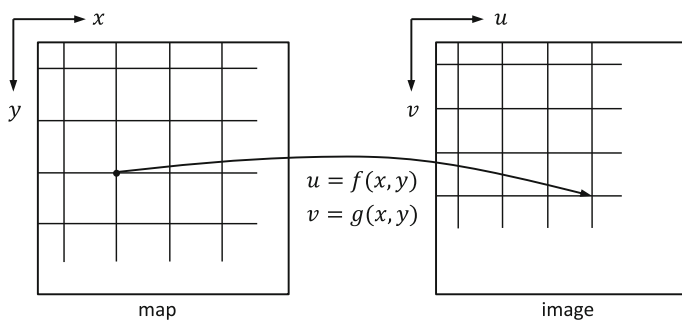
If the coefficients  $a_i$  and  $b_i$  in (2.15) were known then the mapping polynomials could be used to relate any point in the map to its corresponding point in the image, as in the discussion above. At present, however, those coefficients are unknown. Values can be estimated by identifying sets of features on the map that can also be identified on the image. Those features, referred to as *ground control points* (GCPs) or just *control points* (CPs), are well-defined and spatially small. They could be road intersections, street corners, airport runway intersections, sharp bends in rivers, prominent coastline features and the like. Enough are chosen, as pairs on the map and image as depicted in Fig. 2.18, so that the coefficients can be estimated by substituting the coordinates of the control points into the mapping polynomials to yield sets of equations in  $a_i$  and  $b_i$ .

Equations (2.15) show that the minimum number of control points required for second order polynomial mapping is six. Likewise a minimum of three is required for first order mapping and ten for third order mapping. In practice, significantly more than those minimums are chosen and the coefficients are evaluated using least squares estimation. In that manner any control points that contain significant positional errors, either on the map or in the image, will not have an undue influence on the estimated polynomial coefficients.





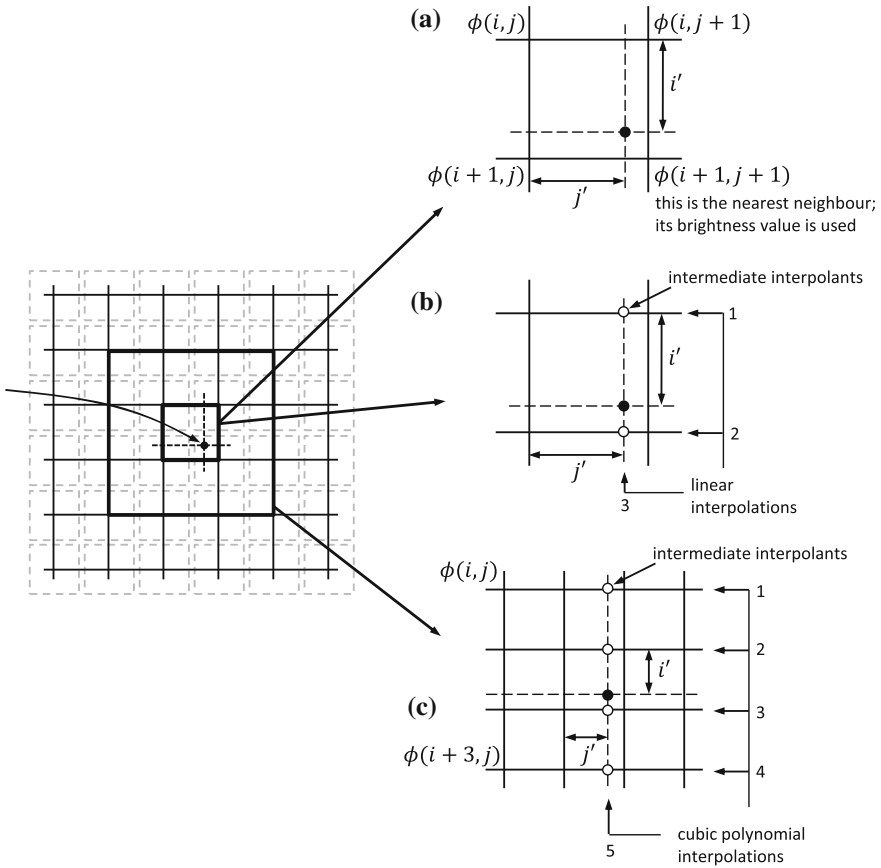
**Fig. 2.18** Map and image coordinate systems, and the concept of ground control points



**Fig. 2.19** Using mapping functions to locate points in the image corresponding to particular display grid (map) positions

### 2.18.2 Building a Geometrically Correct Image

Having specified the mapping polynomials completely by the use of ground control points, the next step is to find points in the image that correspond to each location in the display grid. The spacing of that grid is chosen according to the pixel size required in the corrected image and need not be the same as that of the original, geometrically distorted version. For the moment suppose that the points located in the image correspond exactly to image pixel centres, even though that rarely happens in practice. Then those pixels are simply transferred to the appropriate locations on the display grid to build up the rectified image. That is the case illustrated in Fig. 2.19.



**Fig. 2.20** Determining a pixel brightness value for the display grid by **a** nearest neighbour resampling, **b** bilinear interpolation and **c** cubic convolution interpolation;  $i, j$  etc. are discrete values of  $u, v$

### 2.18.3 Resampling and the Need for Interpolation

As is to be expected, grid centres from the map-defined display grid will not usually project to exact pixel centre locations in the image. This is indicated in Fig. 2.20; some decision now has to be made about what pixel brightness value should be chosen for placement on the new grid.<sup>26</sup> Three principal techniques are used for this purpose.

*Nearest neighbour resampling* simply chooses the actual pixel that has its centre nearest to the point located in the image, as illustrated in Fig. 2.20a. That pixel value is

<sup>26</sup> In some treatments this is referred to as a radiometric transformation; see Le Moigne et al., *loc. cit.*

transferred to the corresponding display grid location. This is the preferred technique if the new image is to be classified because it consists only of original pixel brightnesses, simply rearranged in position to give the correct image geometry. The method is only acceptable when the new and old pixel sizes and spacings are not too different.

*Bilinear interpolation* uses three linear interpolations over the four pixels surrounding the point found in the image projected from a given display grid position. The process is illustrated in Fig. 2.20b. Two linear interpolations are performed along the scan lines as shown to find the interpolants

$$\begin{aligned}\phi(i, j + j') &= j' \phi(i, j + 1) + (1 - j') \phi(i, j) \\ \phi(i + 1, j + j') &= j' \phi(i + 1, j + 1) + (1 - j') \phi(i + 1, j)\end{aligned}$$

where  $\phi$  is pixel brightness and  $(i + i', j + j')$  is the position at which the interpolated value for brightness is required. Position is measured with respect to  $(i, j)$  and, for convenience, assumes a grid spacing of unity in both directions. The final step is to carry out a linear interpolation over  $\phi(i, j + j')$  and  $\phi(i + 1, j + j')$  to give

$$\begin{aligned}\phi(i + i', j + j') &= (1 - i') \{j' \phi(i, j + 1) + (1 - j') \phi(i, j)\} \\ &\quad + i' \{j' \phi(i + 1, j + 1) + (1 - j') \phi(i + 1, j)\}\end{aligned}\quad (2.16)$$

*Cubic convolution interpolation* uses the surrounding sixteen pixels to generate a value for the pixel to be placed on the display grid. Cubic polynomials are fitted along the four lines of four pixels surrounding the point in the image, as shown in Fig. 2.20c to form four interpolants. A fifth cubic polynomial is then fitted through the interpolants to synthesise a brightness value for the corresponding location in the display grid.

The actual form of polynomial that is used for the interpolation is derived from considerations in sampling theory and by constructing a continuous function (an interpolant) from a set of samples.<sup>27</sup> The algorithm that is generally used to perform cubic convolution interpolation is<sup>28</sup>

$$\begin{aligned}\phi(I, j + 1 + j') &= j' \{j' [j' (\phi(I, j + 3) - \phi(I, j + 2) + \phi(I, j + 1) - \phi(I, j))] \\ &\quad + [\phi(I, j + 2) - \phi(I, j + 3) - 2\phi(I, j + 1) + 2\phi(I, j)] \\ &\quad + [\phi(I, j + 2) - \phi(I, j)]\} \\ &\quad + \phi(I, j + 1)\end{aligned}\quad (2.17a)$$

<sup>27</sup> An excellent treatment of the problem has been given by S. Shlien, Geometric correction, registration and resampling of Landsat imagery, *Canadian J. Remote Sensing*, vol. 5, 1979, pp. 74–89. He discusses several possible cubic polynomials that could be used for the interpolation process and demonstrates that the interpolation is a convolution operation.

<sup>28</sup> Based on the choice of interpolation polynomial in T.G. Moik, *Digital Processing of Remotely Sensed Images*, NASA, Washington, 1980.

with  $I = i + n, n = 0, 1, 2, 3$  for the four lines of pixels surrounding the point for which the value is required. Note in Fig. 2.20c we have, for convenience, redefined the address of the  $i, j$  pixel to be one line and column earlier. The four values from (2.17a) are interpolated vertically to give the estimate required

$$\begin{aligned}
 \phi(i + 1 + i', j + 1 + j') = & i' \{ i' [ \phi(i + 3, j + 1 + j') - \phi(i + 2, j + 1 + j') \\
 & + \phi(i + 1, j + 1 + j') - \phi(i, j + 1 + j') ] \\
 & + [ \phi(i + 2, j + 1 + j') - \phi(i + 3, j + 1 + j') \\
 & - 2\phi(i + 1, j + 1 + j') + 2\phi(i, j + 1 + j') ] \\
 & + [ \phi(i + 2, j + 1 + j') - \phi(i, j + 1 + j') ] \} \\
 & + \phi(i + 1, j + 1 + j')
 \end{aligned} \tag{2.17b}$$

Cubic convolution interpolation, or resampling, yields a rectified image that is generally smooth in appearance and is used if the final product is to be analysed visually. However since it gives pixels on the display grid with brightnesses that are interpolated from the original data, it is not recommended if classification is to follow since the new brightness values may be slightly different from the actual radiance values measured by the sensors.

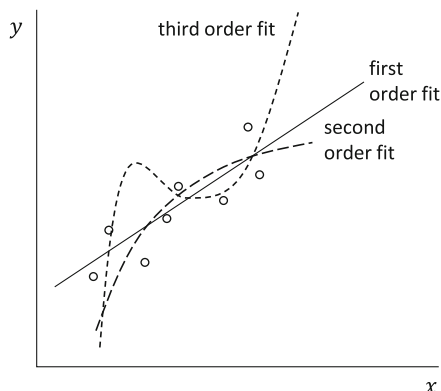
The three interpolation methods just treated are not the only choices, although they are the most common; any effective two dimensional interpolation procedure could be used.<sup>29</sup>

### 2.18.4 The Choice of Control Points

When rectifying an image enough, well-defined, control point pairs must be chosen to ensure that accurate mapping polynomials are generated. Care must also be given to the locations of the points. A general rule is that there should be a distribution of control points around the edges of the image to be corrected, with a scattering of points over the body of the image. That is necessary to ensure that the mapping polynomials are well-behaved over the scene. This concept can be illustrated by considering an example from curve fitting. While the nature of the problem is different, the undesirable effects that can be generated are similar. A set of data points is illustrated in Fig. 2.21 through which first order (linear), second order and third order curves are shown. As the order increases the curves pass closer to the actual points. If it is presumed that the data would have existed for larger values of  $x$ , with much the same trend as apparent in the points plotted, then the linear fit will extrapolate moderately acceptably. In contrast, the cubic curve can deviate markedly from the trend when used as an extrapolator. This is

<sup>29</sup> For other methods see Le Moigne et al., *loc. cit.*

**Fig. 2.21** Illustration from curve fitting to emphasise the potentially poor behaviour of higher order polynomials when used to extrapolate



essentially true in geometric correction of image data: while higher order polynomials may be more accurate in the vicinity of the control points, they can lead to significant errors, and thus distortions, for regions of an image outside the range of the control points. That will be seen in the example of Fig. 2.24.

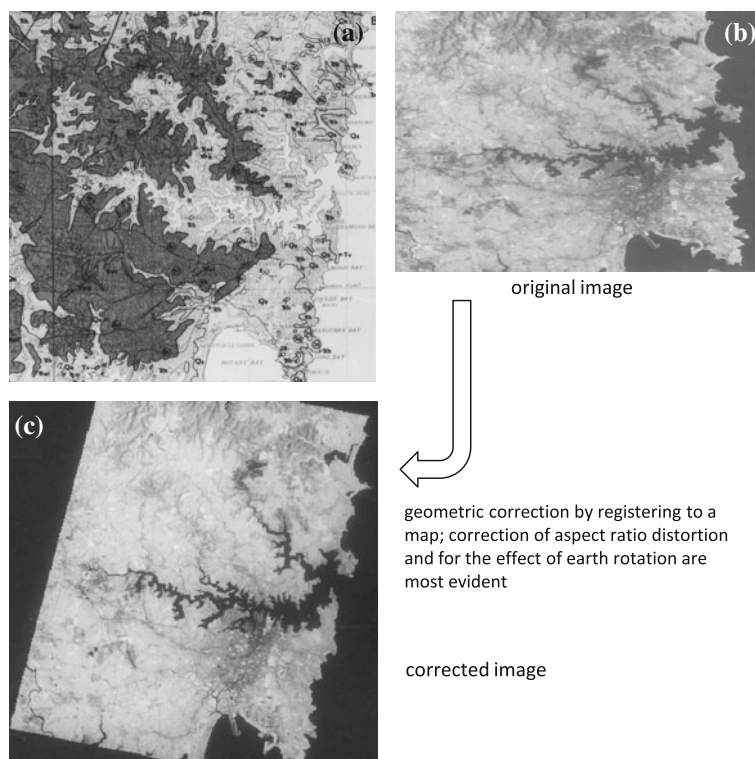
### 2.18.5 Example of Registration to a Map Grid

To illustrate the techniques treated above a small segment of a Landsat MSS image of Sydney, Australia was registered to a map of the region.

It is important that the map has a scale not too different from the scale at which the image data is considered useful. Otherwise the control point pairs may be difficult to establish. In this case a map at 1:250,000 scale was used. The relevant segment is shown reproduced in Fig. 2.22, along with the portion of image to be registered. Comparison of the two demonstrates the geometric distortion of the image. Eleven control points were chosen for the registration, with the coordinates shown in Table 2.1.

Second order mapping polynomials were generated from the set of control points. To test their effectiveness in transferring pixels from the raw image grid to the map display grid, the UTM coordinates of the control points can be computed from their pixel coordinates in the image. They are compared with the actual UTM coordinates and the differences (residuals) calculated in both directions.<sup>30</sup> The root mean square of all the residuals is then computed in both directions (easting and northing) as shown in Table 2.1, giving an overall impression of the accuracy of the mapping process. In this case the control points lead to an average positional error of 56 m in easting and 63 m in northing, which is smaller than a pixel size in equivalent ground metres and thus would be considered acceptable.

<sup>30</sup> This registration exercise was carried out using the Dipix Systems Ltd R-STREAM Software.



**Fig. 2.22** a Map, b segment of a Landsat MSS image exhibiting geometric distortion, and c the result of registering the image to the map, using second order mapping polynomials and cubic convolution resampling, to correct the geometry

At this stage the table can be inspected to see if any individual control point has residuals that are unacceptably high. That could be the result of poor placement; if so, the control point coordinates would be re-entered and the polynomial recalculated. If re-entering the control point leaves the residuals unchanged it may be that there is significant local distortion in that particular region of the image. A choice has to be made then as to whether the control point should be used to give a degree of local correction, that might also influence the remainder of the image, or whether it should be removed and leave that part of the image in error.

Cubic convolution resampling was used in this illustration to generate the image on a  $50 \text{ m} \times 50 \text{ m}$  grid shown in Fig. 2.22c.

Once an image has been registered to a map coordinate system its pixels are addressable in terms of map coordinates rather than pixel and line numbers. Other spatial data types, such as geophysical measurements, can also be registered to the map thus creating a geo-referenced integrated spatial data base of the type used in

**Table 2.1** Control points used in the image to map registration example

GCP	Image pixel	Image line	Map easting actual	Map easting estimate	Map easting residual	Map northing actual	Map northing estimate	Map northing residual
1	1909	1473	432279	432230.1	49.4	836471	836410.1	60.7
2	1950	1625	431288	431418.0	−130.1	822844	822901.4	−56.9
3	1951	1747	428981	428867.9	112.6	812515	812418.2	96.8
4	1959	1851	427164	427196.9	−33.2	803313	803359.4	−46.7
5	1797	1847	417151	417170.3	−18.9	805816	805759.3	57.1
6	1496	1862	397860	397871.6	−11.2	808128	808187.2	−59.6
7	1555	1705	404964	404925.8	38.6	821084	820962.6	121.6
8	1599	1548	411149	411138.5	10.5	833796	833857.3	−61.1
9	1675	1584	415057	415129.0	−72.4	829871	829851.1	19.8
10	1829	1713	422019	421986.6	32.7	816836	816884.5	−48.1
11	1823	1625	423530	423507.8	22.0	824422	824504.8	−83.2
Standard error in easting = 55.92 m								
Standard error in northing = 63.06 m								

a geographic information system. Expressing image pixel addresses in terms of a map coordinate base is referred to as *geocoding*.

**2.19 Mathematical Representation and Correction of Geometric Distortion**

If a particular distortion in image geometry can be represented mathematically then the mapping functions in (2.14) can be specified explicitly. That removes the need to choose arbitrary polynomials as in (2.15) and to use control points to determine the polynomial coefficients. In this section some of the more common distortions are treated from this point of view. Rather than commence with expressions that relate image coordinates  $(u, v)$  to map coordinates  $(x, y)$  it is simpler conceptually to start the other way around, i.e. to model what the true (map) positions of pixels should be, given their positions in an image. This expression can then be inverted, if required, to allow the image to be resampled onto the map grid.

**2.19.1 Aspect Ratio Correction**

The easiest source of geometric error to model is the distortion in aspect caused when the sampling rate across a scan line does not precisely match the IFOV of the sensor. A typical example is that caused by the 56 m ground spacing of the 79 m

$\times 79$  m pixels in the Landsat multispectral scanner. As noted in [Sect. 2.16.3](#) that leads to an image that is too wide for its height by a factor of 1.411. Consequently, to produce a geometrically correct image either the vertical dimension has to be expanded by this amount or the horizontal dimension must be compressed. We consider the former. That requires the pixel axis horizontally to be left unchanged (i.e.  $x = u$ ), but the axis vertically to be scaled (i.e.  $y = 1.411v$ ). These can be expressed conveniently in matrix notation as

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1.411 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \quad (2.18)$$

One way of implementing this correction would be to add extra lines of pixel data to expand the vertical scale which, for the multispectral scanner, could be done by duplicating about four lines in every ten. Alternatively, and more precisely (2.18) can be inverted to give

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0.709 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (2.19)$$

As with the techniques of the previous section, a display grid with coordinates  $(x, y)$  is defined over the map and (2.19) is used to find the corresponding location in the image  $(u, v)$ . The interpolation techniques of [Sect. 2.18.3](#) are then used to generate brightness values for the display grid pixels.

### 2.19.2 Earth Rotation Skew Correction

To correct for the effect of earth rotation it is necessary to implement a shift of pixels to the left with the degree of shift dependent on the particular line of pixels, measured with respect to the top of the image. Their line addresses ( $u$ ) are not affected. Using the results of [Sect. 2.12](#) the corrections are implemented by

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

with  $\alpha = -0.056$  for Sydney, Australia. Again this can be implemented in an approximate sense for multispectral scanner data by making one pixel shift to the left every 17 lines of image data measured down from the top, or alternatively the expression can be inverted to give

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1 & -\alpha \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0.056 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (2.20)$$

which again is used with the interpolation procedures from [Sect. 2.18.3](#) to generate display grid pixels.



### 2.19.3 Image Orientation to North–South

Although not strictly a geometric distortion it is inconvenient to have an image that is correct for most major effects but is not oriented vertically in a north–south direction. It will be recalled for example that Landsat orbits are inclined to the north–south line by about  $9^\circ$ , dependent on latitude. To rotate an image by an angle  $\zeta$  in the counter- or anticlockwise direction (as required in the case of Landsat) it is easily shown that<sup>31</sup>

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\zeta & \sin\zeta \\ -\sin\zeta & \cos\zeta \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

so that

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \cos\zeta & -\sin\zeta \\ \sin\zeta & \cos\zeta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (2.21)$$

### 2.19.4 Correcting Panoramic Effects

The discussion in Sect. 2.14 notes the pixel positional error that results from scanning with a fixed IFOV at a constant angular rate. In terms of map and image coordinates the distortion can be described by

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \tan\theta/\theta & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

where  $\theta$  is the instantaneous scan angle, which in turn is related to  $x$  or  $u$  by  $x = h \tan \theta$ ,  $u = h\theta$ , where  $h$  is altitude. Consequently resampling can be carried out according to

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \theta \cot\theta & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} (h/x)\tan^{-1}(x/h) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (2.22)$$

### 2.19.5 Combining the Corrections

Any exercise in image correction usually requires several distortions to be rectified. Using the techniques in Sect. 2.18 it is assumed that all sources are rectified simultaneously. When employing mathematical modelling, a correction matrix has to be devised for each separate source considered important, as in the preceding sub-sections, and the set of matrices combined. For example if the aspect ratio of a

---

<sup>31</sup> J.D. Foley, A. van Dam, S.K. Feiner and J.F. Hughes, *Computer Graphics: Principles and Practice in C*, 2nd ed., Addison-Wesley, Boston, 1995.

Landsat MSS image is corrected first, followed by correction for the effect of earth rotation, the following single linear transformation can be established for resampling.

$$\begin{aligned} \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1.411 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1.411\alpha \\ 0 & 1.411 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \end{aligned}$$

which, for  $\alpha = -0.056$  at Sydney, gives

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1 & 0.056 \\ 0 & 0.079 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

## 2.20 Image to Image Registration

Many applications in remote sensing require two or more scenes of the same geographical region, acquired at different times, to be processed together. Such a situation arises when changes are of interest, in which case registered images allow a pixel by pixel comparison to be made.

Two images can be registered to each other by registering each to a map coordinate base separately in the manner demonstrated in [Sect. 2.18](#). Alternatively, and particularly if georeferencing is not important, one image can be chosen as a *master*, or reference, to which the other, known as the *slave*, is registered. Again, the techniques of [Sect. 2.18](#) are used. However the coordinates  $(x, y)$  are now the pixel coordinates in the master image rather than the map coordinates. As before  $(u, v)$  are the coordinates of the image to be registered (the slave). A benefit in image to image registration is that only one registration step is required, by comparison to two if both are taken back to a map base. Also, spatial correlation algorithms can be used to assist in accurate co-location of control point pairs, as discussed in [Sect. 2.20.1](#).

### 2.20.1 Refining the Localisation of Control Points

In the great majority of applications control points are chosen manually in both the master and slave images, even though automated techniques are now available.<sup>32</sup> It is frequently of value to check the chosen control point pairs to make sure that they are spatially matched as well as possible. Several means are available for assessing

---

<sup>32</sup> See Le Moigne et al., *loc. cit.*

their correspondence, most of which involve selecting a rectangular sample, or window, of pixels surrounding the designated control point in the slave image, and moving and checking it against the master image, as illustrated in Fig. 2.23.

Because of the spatial properties of the pair of images near the control points the best match should occur when the slave window is located exactly over its counterpart region in the master, in which case the master location of the control point is identified. Obviously it is not necessary to move the slave window over the complete master image since the user knows approximately where the control point should occur in the master. It is only necessary to specify a search region in the neighbourhood of the approximate location.

Control point matching procedures of this sort are usually called sequential similarity detection algorithms (SSDA).<sup>33</sup> There are several means by which the match between the window and search region can be computed. Classically, the two could be correlated.<sup>34</sup> If we denote a pixel in the master image as  $R(i,j)$ , and a pixel in the slave image as  $S(i,j)$  then their correlation over a window of pixels, defined about the control point in the master image, can be expressed

$$m(i,j) = \sum_m \sum_n R(m+i, n+j)S(m,n) \triangleq R * S$$

where  $m,n$  are the pixel coordinates referred to the window. Frequently, this expression is squared and normalised by the autocorrelations over the master and slave to give the similarity measure:

$$SSDA = \frac{(R * S)^2}{(R * R)(S * S)} \quad (2.23)$$

which is in the range (0,1). When the window is best matched to the control point in the master image this measure will be high; for a poor mismatch it will be low.

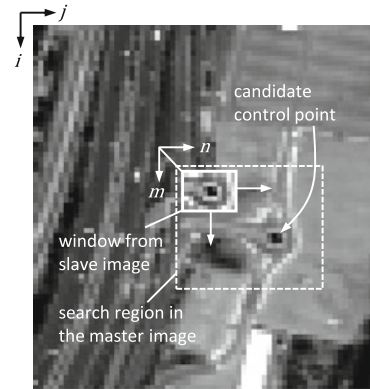
The operation in (2.23) is computationally demanding so simpler matching processes are used in practice. One is based on accumulating the absolute differences in brightness between the master and slave pixels in the window. In principle, when a match is achieved the accumulated difference should be a minimum. This measure is expressed

$$m(i,j) = \sum_m \sum_n |R(m+i, n+j) - S(m,n)| \quad (2.24)$$

<sup>33</sup> See D.I. Barnea and H.F. Silverman, A class of algorithms for fast digital image registration, *IEEE Transactions on Computers*, vol. C-21, no. 2, February 1972, pp. 179–186, and R. Bernstein, Image Geometry and Rectification, in R.N. Colwell, ed., *Manual of Remote Sensing*, 2nd ed., Chap. 21, American Society of Photogrammetry, Falls Church, Virginia, 1983.

<sup>34</sup> See P.E. Anuta, Spatial registration of multispectral and multitemporal digital imagery using fast Fourier transform techniques, *IEEE Transactions on Geoscience Electronics*, vol. GE-8, no. 4, October 1970, pp. 353–368.

**Fig. 2.23** Precisely locating control point pairs by using a window of pixels from the slave image to compare against the master image over a specified search region



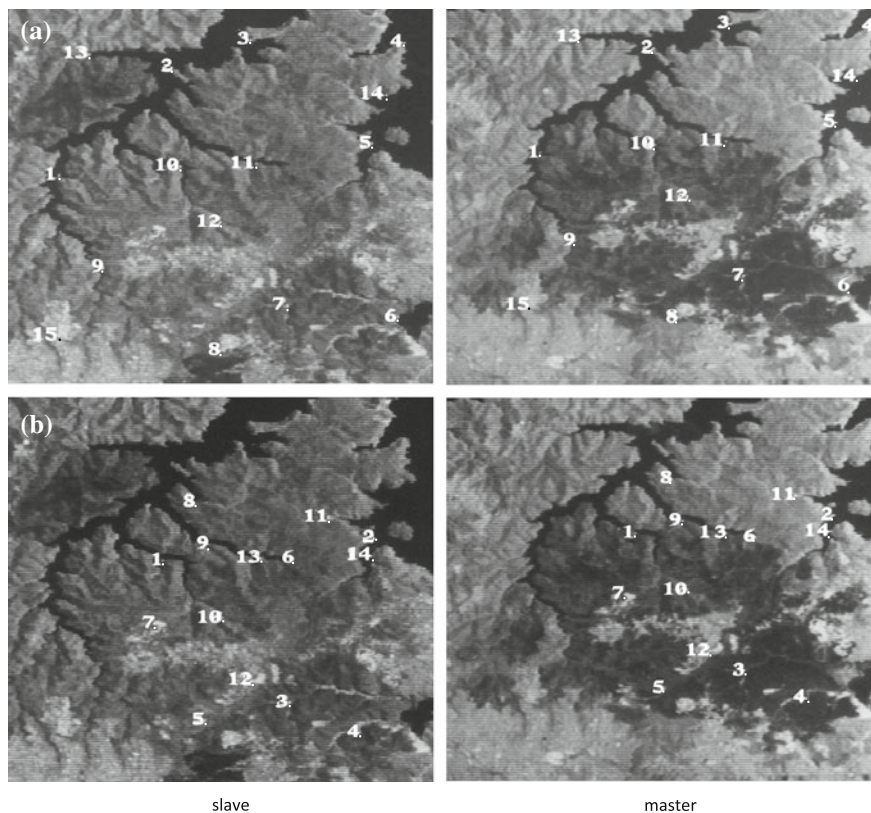
Clearly, measures such as those in (2.23) and (2.24) work well only when the contrast differences over the pixels in the vicinity of the control point are not too different in the master and slave images. If there are significant differences in the distributions of brightness values between the two images, owing to seasonal or noise effects for example, then SSDA techniques will suffer. Alternative procedures include those based on Fourier transforms and the concept of mutual information.<sup>35</sup> Most commercial image processing systems, however, still rely on matching processes such as that in (2.24).

### 2.20.2 Example of Image to Image Registration

To illustrate image to image registration, and particularly to see clearly the effect of control point distribution and the significance of the order of the mapping polynomials to be used for registration, two segments of Landsat Multispectral Scanner infrared image data from the northern suburbs of Sydney were chosen. One was acquired on 29th December, 1979 and was used as the master. The other was acquired on 14th December, 1980 and was used as the slave image. Both are shown in Fig. 2.24; careful inspection shows the differences in image geometry.

Two sets of control points were chosen. In one, the points are distributed as nearly as possible in a uniform manner around the edge of the image segment as shown in Fig. 2.24a, with some points located across the centre of the image. This set would be expected to give reasonable registration of the images. The second set of control points was chosen injudiciously, closely grouped around one particular region, to illustrate the resampling errors that can occur. They are shown in Fig. 2.24b. In both cases the control point pairs were co-located with the assistance of the sequential similarity detection algorithm in (2.24). This worked well

<sup>35</sup> See Le Moigne et al., *loc. cit.*

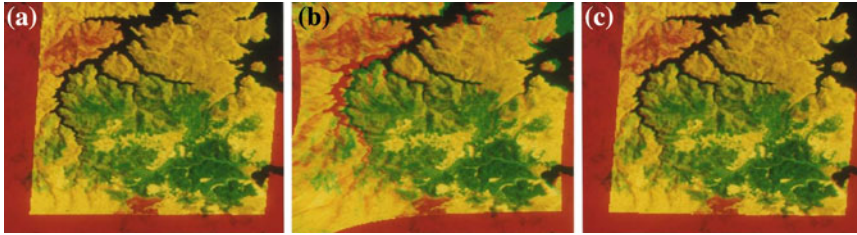


**Fig. 2.24** Control points used in the image to image registration example: **a** good distribution **b** poor distribution

particularly for those control points around the coastal and river regions where the similarity between the images is unmistakable. To minimise tidal influences on the location of control points, those on water boundaries were chosen as near as possible to be on headlands, and never at the ends of inlets.

For both sets of control points third order mapping polynomials were used, along with cubic convolution resampling. As expected the first set of points led to an acceptable registration of the images whereas the second set gave a good registration in the immediate neighbourhood of the control points but beyond that neighbourhood produced gross distortion.

The adequacy of the registration process can be assessed visually if the master and resampled slave images are superimposed in different colours. Figures 2.25a,b show the master image in red with the resampled slave image superimposed in green. Where good registration has been achieved the result is yellow, with the exception of regions of gross dissimilarity in pixel brightness—in this case associated with fire burns. Misregistration shows quite graphically as a red-green



**Fig. 2.25** **a** Registration of 1980 image (*green*) with 1979 image (*red*) using the control points of Fig. 2.24a, and third order mapping polynomials **b** registration of 1980 image (*green*) with 1979 image (*red*) using the control points of Fig. 2.24b, and third order mapping polynomials **c** registration of 1980 image (*green*) with 1979 image (*red*) using the control points of Fig. 2.24b, and first order mapping polynomials

separation. This is particularly noticeable in Fig. 2.25b where the poor extrapolation obtained with third order mapping is demonstrated.

The exercise using the poor set of control points in Fig. 2.24b was repeated. However this time first order mapping polynomials were used. While they will not remove non-linear differences between the images, and will give poorer matches at the control points themselves, they are well behaved in extrapolation beyond the vicinity of the control points and lead to an acceptable registration as seen in Fig. 2.25c.

## 2.21 Other Image Geometry Operations

While the techniques of the previous sections have been devised for treating errors in image geometry, and for registering sets of images, they can also be used for performing intentional changes to image geometry. Image rotation and scale changing are chosen here as illustrations.

### 2.21.1 Image Rotation

Rotation of an image by an angle about the pixel grid can be useful for a number of applications. Most often it is used to align the pixel grid, and thus the image, to a north–south orientation as treated in Sect. 2.19.3. However the transformation in (2.21) is perfectly general and can be used to rotate an image in an anticlockwise sense by any specified angle  $\zeta$ .

### 2.21.2 Scale Changing and Zooming

The scales of an image in both the vertical and horizontal directions can be altered by the transformation

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix}$$

where  $a$  and  $b$  are the desired scaling factors. To resample the scaled image onto the display grid we use the inverse operation to locate pixel positions in the original image corresponding to each display grid position, viz.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1/a & 0 \\ 0 & 1/b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Interpolation is used to establish the actual pixel brightnesses to use, since  $u, v$  will not normally fall on exact pixel locations.

Frequently  $a = b$  so that the image is simply magnified. This is called *zooming*. If the nearest neighbour interpolation procedure is used in the resampling process the zoom implemented is said to occur by *pixel replication* and the image will look progressively blocky for larger zoom factors. If cubic convolution interpolation is used there will be a change in magnification but the image will not take on the blocky appearance. Often this process is called *interpolative zoom*.

## 2.22 Bibliography on Correcting and Registering Images

A good general discussion on the effects of the atmosphere on the passage of radiation in the range of wavelengths important to optical remote sensing will be found in

P.N. Slater, *Remote Sensing: Optics and Optical Systems*, Addison-Wesley, Reading, Mass., 1980.

An introduction to the radiometric distortion problem facing high spectral resolution imagery, such as that produced by imaging spectrometers, is given, along with correction techniques, in

B.C. Gao, K. B. Heidebrecht and A.F.H. Goetz, Derivation of scaled surface reflectance from AVIRIS data, *Remote Sensing of Environment*, vol. 44, 1993, pp. 165–178 and

B. C Gao, M.J. Montes, C.O. Davis and A.F.H. Goetz, Atmospheric correction algorithms for hyperspectral remote sensing data of land and oceans, *Remote Sensing of Environment*, Supplement 1, Imaging Spectroscopy Special Issue, vol. 113, 2009, pp. S17–S24, while

F.A. Kruse, Comparison of ATREM, ACORN and FLAASH atmospheric corrections using low-altitude AVIRIS data of Boulder, Colorado, *Proc. 13th JPL Airborne Geoscience Workshop*, Pasadena, CA, 2004

gives a very helpful comparison of the more common correction procedures.

Techniques for correcting errors in image geometry are discussed in many standard image processing treatments in remote sensing. More recently, the range of operational procedures and new research directions in geometric correction and image registration are covered in

J. Le Moigne, N. S. Netanyahu and R. D. Eastman, eds., *Image Registration for Remote Sensing*, Cambridge University Press, Cambridge, 2011.

A treatment of experimental fully automated image registration, including a review of available techniques, is given in

H. Gonçalves, L. Corte-Real and J.A. Gonçalves, Automatic image registration through image segmentation and SIFT, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 7, July 2011, pp. 2589–2600.

Some earlier treatments should not be overlooked as they provide good insight into the problem of correcting geometry and are still relevant. They include

S. Shlien, Geometric correction, registration and resampling of Landsat imagery, *Canadian J. Remote Sensing*, vol. 5, 1979, pp. 74–89, and

F. Orti, Optimal distribution of control points to minimise Landsat registration errors, *Photogrammetric Engineering and Remote Sensing*, vol. 47, 1980, pp. 101–110.

Finally, books on computer graphics also contain very good material on image geometry correction and transformation, perhaps one of the most notable being

J.D. Foley, A. van Dam, S.K. Feiner and J.F. Hughes, *Computer Graphics: Principles and Practice in C*, 2nd ed., Addison-Wesley, Boston, 1995.

## 2.23 Problems

- 2.1 (a) Consider a region on the ground consisting of a square grid. For simplicity suppose the grid lines are 79 m in width and the grid spacing is 790 m. Sketch how the region would appear in Landsat multispectral scanner imagery before any geometric correction has been applied. Include only the effect of earth rotation and the effect of 56 m horizontal spacing of the 79 m  $\times$  79 m ground resolution elements.
- (b) Develop a pair of linear mapping polynomials that will correct the image in (a). Assume the “lines” on the ground have a brightness of 100 and that the background brightness is 20. Resample onto a 50 m grid and use nearest neighbour interpolation. You will not want to compute all the resampled pixels unless a computer is used for the exercise. Instead, simply consider some significant pixels in the resampling to illustrate the accuracy of the geometric correction.
- 2.2 A sample of pixels from each of three cover types present in the Landsat MSS scene of Sydney, Australia, acquired on 14th December, 1980 is given in Table 2.2a. Only the brightnesses (digital numbers) in the visible red band (0.6–0.7  $\mu\text{m}$ ) and the second of the infrared bands (0.8–1.1  $\mu\text{m}$ ) are given.



**Table 2.2** Digital numbers for a set of pixels from three cover types; note band 5 covers the wavelength range 0.6–0.7 μm and band 7 covers the range 0.8–1.1 μm

(a) Landsat MSS image of Sydney 14th December, 1980

Water		Vegetation		Soil	
Band 5	Band 7	Band 5	Band 7	Band 5	Band 7
20	11	60	142	74	66
23	7	53	130	103	82
21	8	63	140	98	78
21	7	52	126	111	86
22	7	34	92	84	67
19	3	38	120	76	67
17	1	38	151	72	67
20	4	38	111	98	71
24	8	31	81	99	80
19	4	50	158	108	71

(b) Landsat MSS image of Sydney 8th June, 1980

Water		Vegetation		Soil	
Band 5	Band 7	Band 5	Band 7	Band 5	Band 7
11	2	19	41	43	27
13	5	24	45	43	34
13	2	20	44	40	30
11	1	22	30	27	19
9	1	15	22	34	23
14	4	14	26	36	26
13	4	21	27	34	27
15	5	17	38	70	50
12	4	24	37	37	30
15	4	20	27	44	30

For this image Forster<sup>36</sup> has computed the following relations between reflectance ( $R$ ) and digital number ( $DN$ ), where the subscript 7 refers to the infrared data and the subscript 5 refers to the visible red data:

$$R_5 = 0.44 DN_5 + 0.5$$

$$R_7 = 1.18 DN_7 + 0.9$$

Table 2.2b shows samples of MSS digital numbers for a second scene of Sydney recorded on 8th June, 1980. For this image Forster has determined

$$R_5 = 3.64 DN_5 - 1.6$$

$$R_7 = 1.52 DN_7 - 2.6$$

<sup>36</sup> See B.C. Forster, *loc. cit.*

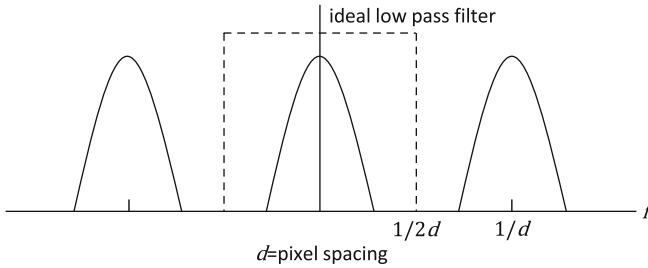
Compute the mean digital count value for each cover type in each scene and plot these, along with bars at  $\pm 1$  standard deviation, in a spectral domain that has the infrared values on the ordinate and the visible red values along the abscissa. Now produce the same plots after converting the data to reflectances. Comment on the effect that correction of the raw digital numbers to reflectance data, in which atmospheric effects have been removed, has on the apparent separation of the three cover types in the spectral domain.

- 2.3 Aircraft line scanners frequently use a rotating mirror that sweeps out lines of data at right angles to the fuselage to acquire imagery. In the absence of a cross wind scanning will be orthogonal to the aircraft ground track. Often scanning is carried out in the presence of a cross wind. The aircraft fuselage then maintains an angle to the ground track so that scanning is no longer orthogonal to the effective forward motion, leading to a distortion referred to as *crabbing*. Discuss the nature of this distortion when the image pixels are displayed on a square grid. Remember to take account of the finite time required to scan across a line of pixels.

Push broom scanners are also used on aircraft. What is the nature of the geometric distortion incurred with those sensors in the presence of a cross wind?

Complete frames of pixels can also be captured from aircraft platforms, effectively using digital cameras. What geometric distortion would be incurred with such a sensor in the presence of a cross wind?

- 2.4 Compute the skew distortion resulting from earth rotation in the case of Landsat 7 and SPOT.
- 2.5 For a particular application suppose it was necessary to apply geometric correction procedures to an image prior to classification. (See [Chap. 3](#) for an overview of classification). What interpolation technique would you prefer to use in the resampling process? Why?
- 2.6 Destriping Landsat multispectral scanner images is often performed by computing six modulo-6 line histograms and then either (i) matching all six to a standard histogram or (ii) choosing one of the six as a reference and matching the other five to it. Which method is to be preferred if the image is to be analysed by photointerpretation or by classification?
- 2.7 In a particular problem you have to register five images to a map. Would you register each image to the map separately, register one image to the map and then the other four images to that one, or image 1 to the map, image 2 to image 1, image 3 to image 2 etc.?
- 2.8 (This requires a background in digital signal processing and sampling theory—see [Chap. 7](#)). Remote sensing digital images are uniform two dimensional samples of the ground scene. One line of image data is a regular sequence of samples. The spatial frequency spectrum of a line data will therefore be periodic as depicted in [Fig. 2.26](#); the data can be recovered by low pass filtering the spectrum, using the ideal filter indicated in the figure.



**Fig. 2.26** Idealised spatial frequency spectrum of line samples (pixels)

Multiplication of the spectrum by this ideal filter is equivalent to convolving the original line of samples by the inverse Fourier transform of the filter function. From the theory of the Fourier transform, the inverse of the filter function is

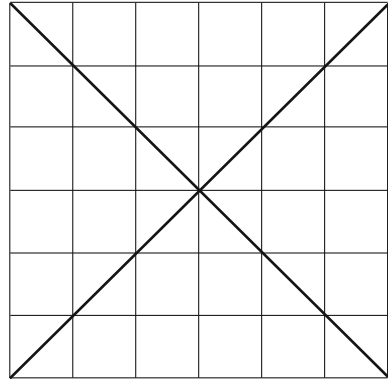
$$s(x) = \frac{2d}{\pi} \frac{\sin x}{x}$$

with  $x = \zeta/2d$  in which  $\zeta$  is a spatial variable along lines of data, and  $d$  is the inter-pixel spacing.  $s(x)$  is known generally as an interpolating function.

Determine some cubic polynomial approximations to this function. These could be determined from a simple Taylor series expansion or could be derived from cubic splines.<sup>37</sup>

- 2.9 A scanner has been designed for aircraft operation. It has a field of view (FOV) of  $\pm 35^\circ$  about nadir and an instantaneous field of view (IFOV) of 2 mrad. The sensor is designed to operate at a flying height of 1000 m.
  - (i) Determine the pixel size, in metres, at nadir.
  - (ii) Determine the pixel size at the edge of a swath compared with that at nadir.
  - (iii) Discuss the nature of the distortion in image geometry encountered if the pixels across a scan line are displayed on uniform pixel centres.
- 2.10 Determine the maximum angle of the field of view (FOV) for an airborne optical sensor with a constant instantaneous field of view (IFOV), so that the pixel dimension along the scan line at the extremes is less than 1.5 times that at nadir (ignore the earth curvature effect).
- 2.11 Consider the panoramic along scan line distortion of an airborne optical remote sensing system with a constant instantaneous field of view (IFOV); sketch the image formed for the ground scene shown in Fig. 2.27 and explain why it appears as you have sketched it.

<sup>37</sup> For some examples see Shlien, *loc. cit.*

**Fig. 2.27** Ground scene

- 2.12 Compare the flat field, empirical line and log residuals methods for radiometric correction from the point of view of ease of use compared with the nature of the result.
- 2.13 Following the development in [Sect. 2.19](#) derive a single matrix that describes the transformation from map coordinates to image coordinates for each of the following instruments, taking into account only earth rotation and orbital inclination at the equator.

SPOT HRG

AQUA MODIS

Ikonos

- 2.14 Plot a graph of pixel size across the swath, as a function of pixel size at nadir, for look angles out to  $70^\circ$ ; locate on the graph individual results for SPOT HRG, SPOT Vegetation, TERRA MODIS and EO-1 ALI.
- 2.15 Suppose a geostationary satellite carries a simple imager in the form of a digital camera. If the camera's field of view were such that it imaged a square area on the earth's surface with 500 km sides discuss any distortions that might be present in the image and how they would appear in the final image product.