1 General image characteristics, data acquisition and image reconstruction

1.1 Introduction

A clinician making a diagnosis based on medical images looks for a number of different types of indication. These could be changes in shape, for example enlargement or shrinkage of a particular structure, changes in image intensity within that structure compared to normal tissue and/or the appearance of features such as lesions which are normally not seen. A full diagnosis may be based upon information from several different imaging modalities, which can be correlative or additive in terms of their information content.

Every year there are significant engineering advances which lead to improvements in the instrumentation in each of the medical imaging modalities covered in this book. One must be able to assess in a quantitative manner the improvements that are made by such designs. These quantitative measures should also be directly related to the parameters which are important to a clinician for diagnosis. The three most important of these criteria are the spatial resolution, signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR). For example, Figure 1.1(a) shows a magnetic resonance image with two very small white-matter lesions indicated by the arrows. The spatial resolution in this image is high enough to be able to detect and resolve the two lesions. If the spatial resolution were to have been four times worse, as shown in Figure 1.1(b), then only the larger of the two lesions is now visible. If the image SNR were four times lower, illustrated in Figure 1.1(c), then only the brighter of the two lesions is, barely, visible. Finally, if the CNR between the lesions and the surrounding white matter is reduced, as shown in Figure 1.1(d), then neither lesion can be discerned.

Although one would ideally acquire images with the highest possible SNR, CNR and spatial resolution, there are often trade-offs between the three parameters in terms of both instrument design and data acquisition techniques, and careful choices must be made for the best diagnosis. This chapter covers the quantitative aspects of assessing image quality, some of the trade-offs between SNR, CNR and spatial resolution, and how measured data can be digitized, filtered and stored. At the end of this chapter, the two essential algorithms for reconstruction of

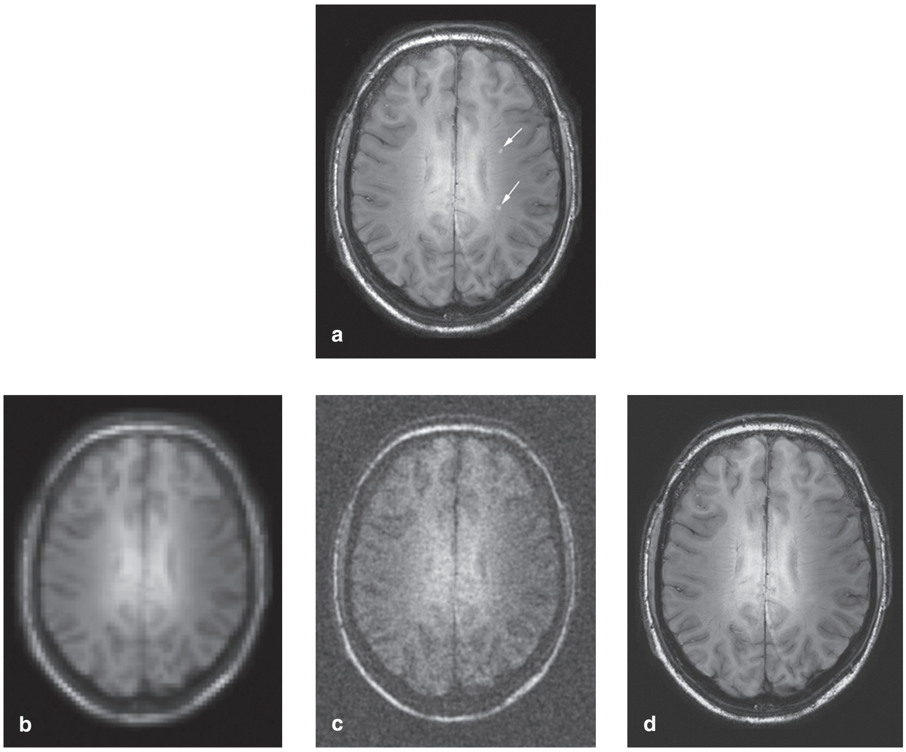

Figure 1.1

(a) MR image showing two small white-matter lesions indicated by the arrows. Corresponding images acquired with (b) four times poorer spatial resolution, (c) four times lower SNR, and (d) a reduced CNR between the lesions and the surrounding healthy tissue. The arrows point to lesions that can be detected.

medical images, namely the Fourier transform and filtered backprojection, are introduced.

1.2 Specificity, sensitivity and the receiver operating characteristic (ROC) curve

The accuracy of clinical diagnoses depends critically upon image quality, the higher the quality the more accurate the diagnosis. Improvements in imaging techniques and instrumentation have revolutionized early diagnosis and treatment of a number of different pathological conditions. Each new imaging technique or change in instrumentation must be carefully assessed in terms of its effect on diagnostic accuracy. For example, although the change from planar X-ray film to digital radiography clearly has many practical advantages in terms of data storage and mobility, it would not have been implemented clinically had the diagnostic quality of the scans decreased. Quantitative assessment of diagnostic quality is usually reported in terms of specificity and sensitivity, as described in the example below.

Consider an imaging study to determine whether a group of middle-aged patients has an early indication of multiple sclerosis. It is known that this disease is characterized by the presence of white matter lesions in the brain. However, it is also known that healthy people develop similar types of lesion as they age, but that the number of lesions is not as high as for multiple sclerosis cases. When analyzing the images from a particular patient there are four possible outcomes for the radiologist: a true positive (where the first term 'true' refers to a correct diagnosis and the second term 'positive' to the patient having multiple sclerosis), a true negative, a false positive or a false negative. The four possibilities can be recorded in either tabular or graphical format, as shown in Figure 1.2. The receiver operating characteristic (ROC) curve plots the number of true positives on the vertical axis vs. the number of false positives on the horizontal axis, as shown on the right of Figure 1.2. What criterion does the radiologist use to make his/her diagnosis? In this simple example assume that the radiologist simply counts the number of lesions detectable in the image. The relative number of true positives, true negatives, false positives and false negatives depends upon the particular number of lesions that the radiologist decides upon as being the threshold for diagnosing a patient with multiple sclerosis. If this threshold number is very high, for example 1000, then there will be no false positives, but no true positives either. As the threshold number is reduced then the number of true positives will increase at a greater rate than the false positives, providing that the images are giving an accurate count of the number of lesions actually present. As the criterion for the number of lesions is reduced further, then the numbers of false positives and true positives increase at a more equal rate. Finally, if the criterion is dropped to a very small number, then the number of false positives increases much faster than the true positives. The net effect is to produce a curve shown in Figure 1.2.

Three measures are commonly reported in ROC analysis:

(i) *accuracy* is the number of correct diagnoses divided by the total number of diagnoses;
(ii) *sensitivity* is the number of true positives divided by the sum of the true positives and false negatives; and
(iii) *specificity* is the number of true negatives divided by the sum of the number of true negatives and false positives.

	Actual situation	
Diagnosis	*Multiple sclerosis*	*Healthy*
Multiple sclerosis	True positive	False positive
Healthy	False negative	True negative

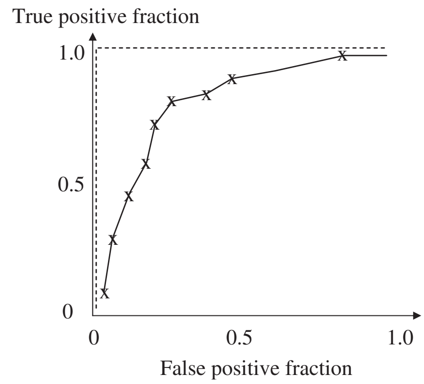

Figure 1.2

The receiver operating characteristic curve. (left) The 2 × 2 table showing the four possible outcomes of clinical diagnosis. (right) A real ROC curve (solid line), with the ideal curve also shown (dotted line).

The aim of clinical diagnosis is to maximize each of the three numbers, with an ideal value of 100% for all three. This is equivalent to a point on the ROC curve given by a true positive fraction of 1, and a false positive fraction of 0. The corresponding ROC curve is shown as the dotted line in Figure 1.2. The closer the actual ROC curve gets to this ideal curve the better, and the integral under the ROC curve gives a quantitative measure of the quality of the diagnostic procedure.

A number of factors can influence the shape of the ROC curve, but from the point-of-view of the medical imaging technology, the relevant question is: ‘what fraction of true lesions is detected using the particular imaging technique?’. The more lesions that are missing on the image, then intuitively the poorer the resulting diagnosis. ‘Missing lesions’ may occur, referring to Figure 1.1, due to poor SNR, CNR or spatial resolution. In turn, these will lead to a decreased percentage accuracy, sensitivity and specificity of the diagnostic procedure.

Example 1.1 Suppose that a criterion used for diagnosis is, in fact, completely unrelated to the actual medical condition, e.g. as a trivial example, trying to diagnose cardiac disease by counting the number of lesions in the brain. Draw the ROC curve for this particular situation.

Solution Since the criterion used for diagnosis is independent of the actual condition, effectively we have an exactly equal chance of a true positive or a false positive, irrespective of the number of lesions found in the brain. Therefore, the ROC curve is a straight line at an angle of 45° to the main axes, as shown below.

1.3 Spatial resolution

The spatial resolution of an imaging system is related to the smallest feature that can be visualized or, more specifically, the smallest distance between two features such that the features can be individually resolved rather than appearing as one larger shape. The two most common measures in the spatial domain are the line spread function (LSF) and point spread function (PSF). These measures can be represented by an equivalent modulation transfer function (MTF) in the spatial frequency domain. The concept of spatial frequency is very useful in characterizing spatial resolution, and is explained in the following section.

1.3.1 Spatial frequencies

One familiar example of spatial frequencies is a standard optician's test. In one test, patients are asked to look at a series of black lines on a white background, and then to tell the optician if they can resolve the lines when a series of lenses with different strengths are used. As shown in Figure 1.3, the lines are of different thickness and separation. The spatial frequency of a particular grid of lines is measured as the number of lines/mm, for example 5 mm^{-1} for lines spaced 200 μm apart. The closer together are the lines, the higher is the spatial frequency, and the better the spatial resolution of the image system needs to be to resolve each individual line.

For each medical imaging modality, a number of factors affect the resolving power. One can simplify the analysis by considering just two general components: first the instrumentation used to form the image, and second the quantity of data that is acquired i.e. the image data matrix size. To take the everyday example of a digital camera, the lens and electronics associated with the charge-coupled device (CCD) detector form the instrumentation, and the number of pixels (megapixels) of the camera dictates the amount of data that is acquired. The relative contribution of each component to the overall spatial resolution is important in an optimal

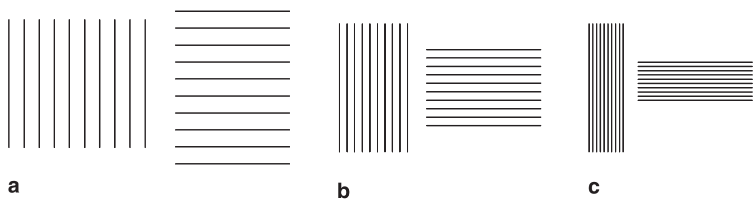

Figure 1.3

Grid patterns with increasing spatial frequencies going from (a) to (c).

engineering design. There is no advantage in terms of image quality in increasing the CCD matrix size from 10 megapixels to 20 megapixels if the characteristics of the lens are poor, e.g. it is not well focused, produces blur, or has chromatic aberration. Similarly, if the lens is extremely well-made, then it would be sub-optimal to have a CCD with only 1 megapixel capability, and image quality would be improved by being able to acquire a much greater number of pixels.

1.3.2 The line spread function

The simplest method to measure the spatial resolution of an imaging system is to perform the equivalent of the optician's test. A single thin line or set of lines is imaged, with the relevant structure made of the appropriate material for each different imaging modality. Examples include a strip of lead for an X-ray scan, a thin tube of radioactivity for nuclear medicine imaging, or a thin wire embedded in gel for an ultrasound scan. Since the imaging system is not perfect, it introduces some degree of blurring into the image, and so the line in the image does not appear as sharp as its actual physical shape. The degree of blurring can be represented mathematically by a line-spread function (LSF), which is illustrated in Figure 1.4.

The LSF of an imaging system is estimated by measuring a one-dimensional projection, as shown in Figure 1.4, with y defined as the horizontal direction. The width of the LSF is usually defined by a parameter known as the full-width-at-half-maximum (FWHM). As the name suggests, this parameter is the width of the particular function at a point which is one-half the maximum value of the vertical axis. From a practical point-of-view, if two small structures in the body are

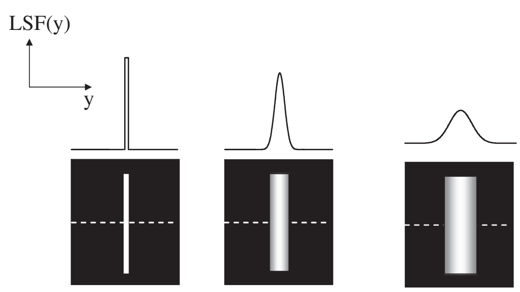

Figure 1.4

The concept of the line-spread function. A thin object is imaged using three different imaging systems. The system on the left has the sharpest LSF, as defined by the one-dimensional projection measured along the dotted line and shown above each image. The system in the middle produces a more blurred image, and has a broader LSF, with the system on the right producing the most blurred image with the broadest LSF.

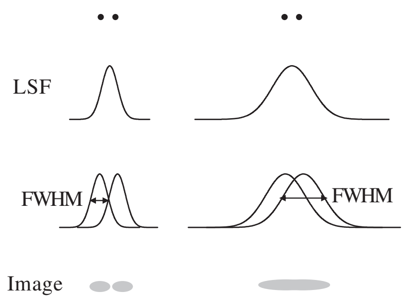

Figure 1.5

Imaging results produced by two different systems with a relatively narrow (left) and broad (right) LSF. In the case on the left, two small structures within the body (top) have a separation which is slightly greater than the FWHM of the LSF, and so the resulting image shows the two different structures. In the case on the right, the FWHM of the LSF is greater than the separation of the structures, and so the image appears as one large structure.

separated by a distance greater than the FWHM of the LSF, then they can be resolved as separate structures as opposed to one larger structure, as shown in Figure 1.5.

The LSF for many imaging techniques is well-approximated by a Gaussian function, defined as:

$$\mathrm{LSF}(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-y_0)^2}{2\sigma^2}\right), \tag{1.1}$$

where σ is the standard deviation of the distribution, and y_0 is the centre of the function. The FWHM of a Gaussian function is given by:

$$\mathrm{FWHM} = \left(2\sqrt{2\ln 2}\right)\sigma \cong 2.36\sigma. \tag{1.2}$$

Therefore, if the physical separation between two structures is greater than 2.36 times the standard deviation of the Gaussian function defining the LSF of the imaging system, then the two structures can be resolved.

1.3.3 The point spread function

The LSF describes system performance in one dimension. However, in some imaging modalities, for example nuclear medicine, the spatial resolution becomes poorer the deeper the location within the patient from which the signal is received. Therefore, a full description of the spatial resolution of an imaging system requires

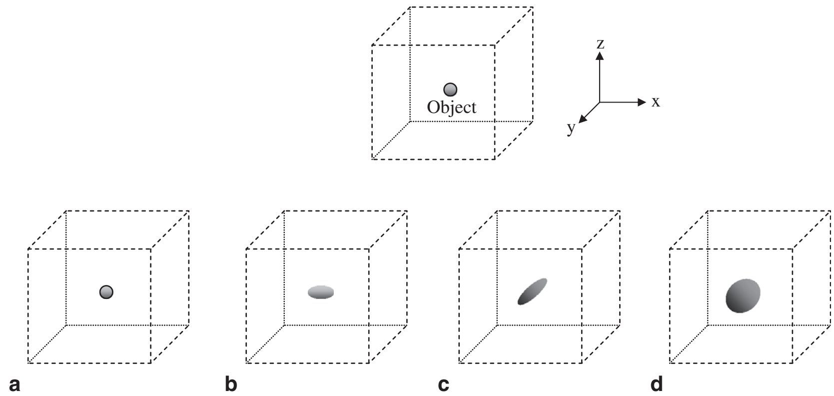

Figure 1.6

(top) Small point object being imaged. (a)-(d) Images produced with different point spread functions. (a) A sharp PSF in all three dimensions. (b) A PSF which is significantly broader in x than in y or z. (c) A PSF which is broadest in the y-dimension. (d) A PSF which is broad in all three dimensions.

a three-dimensional formulation: the three-dimensional equivalent of the LSF is termed the point spread function (PSF). As the name suggests, the PSF describes the image acquired from a very small 'point source', for example a small sphere of water for MRI. Examples of spatially symmetric and asymmetric PSFs are shown in Figure 1.6.

Mathematically, the three-dimensional image (I) and object (O) are related by:

$$I(x, y, z) = O(x, y, z) * h(x, y, z), \tag{1.3}$$

where * represents a convolution, and h(x,y,z) is the three-dimensional PSF. In a perfect imaging system, the PSF would be a delta function in all three dimensions, in which case the image would be a perfect representation of the object. In practice, the overall PSF of a given imaging system is a combination of detection instrumentation and data sampling (covered in Section 1.7), and can be calculated by the convolution of all of the individual components.

1.3.4 The modulation transfer function

The most commonly used measure of the spatial resolution of an imaging system is the modulation transfer function (MTF). This measures the response of a system to both low and high spatial frequencies. An ideal system has a constant MTF for all spatial frequencies, i.e. it exactly reproduces both the fine structure (high spatial frequencies) and areas of relatively uniform signal intensity (low spatial frequencies). In practice, as seen previously, imaging systems have a finite spatial resolution, and the high spatial frequencies must at some value start to be attenuated: the greater the attenuation the poorer the spatial resolution. Mathematically, the MTF is given by the Fourier transform of the PSF:

$$\mathrm{MTF}\left(k_x, k_y, k_z\right) = F\{\mathrm{PSF}(x, y, z)\}, \qquad (1.4)$$

where k_x, k_y and k_z are the spatial frequencies measured in lines/mm corresponding to the x,y and z spatial dimensions measured in mm. Properties of the Fourier transform are summarized in Section 1.9. The relationship between a one-dimensional (for simplicity) MTF and PSF is shown in Figure 1.7. The ideal MTF, which is independent of spatial frequency, corresponds to a PSF which is a delta function. The broader is the PSF the narrower the MTF, and the greater the degree to which the high spatial frequency information is lost.

Since the PSF and MTF are related by the Fourier transform, and a convolution in one domain is equivalent to multiplication in the other (Section 1.9.2), the overall MTF of the imaging system can be calculated by multiplying together the effects of all of the contributing components.

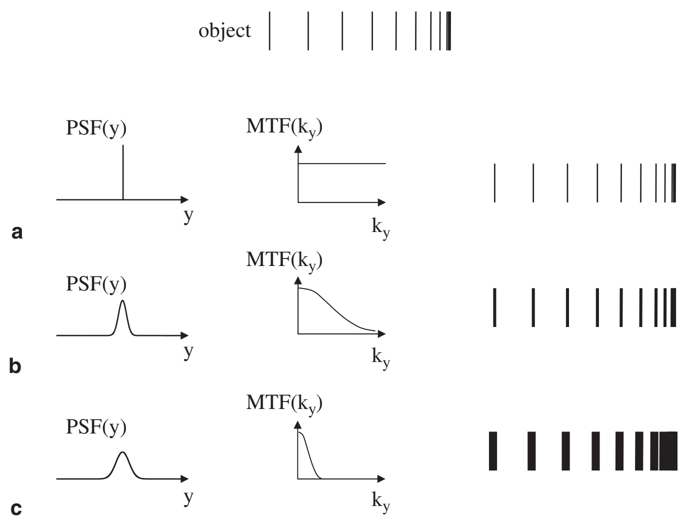

Figure 1.7
(top) The object being imaged corresponds to a set of lines with increasing spatial frequency from left-to-right. (a) An ideal PSF and the corresponding MTF produce an image which is a perfect representation of the object. (b) A slightly broader PSF produces an MTF which loses the very high spatial frequency information, and the resulting image is blurred. (c) The broadest PSF corresponds to the narrowest MTF, and the greatest loss of high spatial frequency information.

1.4 Signal-to-noise ratio

In all measured or recorded signals there is some contribution from noise. Crackle over the radio or on a mobile phone is perhaps the most familiar phenomenon. Noise refers to any signal that is recorded, but which is not related to the actual signal that one is trying to measure (note that this does not include image artifacts, which are considered separately in Section 1.8). In the simplest cases, noise can be considered as a random signal which is superimposed on top of the real signal. Since it is random, the mean value is zero which gives no indication of the noise level, and so the quantitative measure of the noise level is conventionally the standard deviation of the noise. It is important in designing medical imaging instrumentation that the recorded signal is as large as possible in order to get the highest signal-to-noise ratio (SNR). An example of the effects of noise on image quality is shown in Figure 1.8. As the noise level increases, the information content and diagnostic utility of the image are reduced significantly.

The factors that affect the SNR for each imaging modality are described in detail in the relevant sections of each chapter. However, two general cases are summarized here. If the noise is truly random, as in MRI, then the image SNR can be increased by repeating a scan a number of times and then adding the scans together. The true signal is the same for every scan, and so adds up 'coherently': for N co-added scans the total signal is N times that of a single scan. However, the noise at each pixel is random, and basic signal theory determines that the standard deviation of a random variable increases only as the square root of the number of co-added scans. Therefore, the overall SNR increases as the square root of the number of scans. An example from MRI, in which such 'signal averaging' is commonly used, is shown in Figure 1.9. The trade-off in signal averaging is the additional time required for data acquisition which means that signal averaging cannot be used, for example, in dynamic scanning situations.

In ultrasonic imaging the situation is more complicated since the major noise contribution from speckle is coherent, and so signal averaging does not increase the SNR. However, if images are acquired with the transducer oriented at different angles with respect to the patient, a technique known as compound imaging (covered in Section 4.8.4), then the speckle in different images is only partially coherent. Averaging of the images, therefore, gives an increase in the SNR, but by a factor less than the square root of the number of images.

In the second general case, as discussed in detail in Chapters 2 and 3, the SNR in X-ray and nuclear medicine is proportional to the square root of the number of X-rays and γ-rays, respectively, that are detected. This number depends upon many factors including the output dose of the X-ray tube or the amount of