CHAPTER 12

COMPUTERS IN NUCLEAR MEDICINE

J.A. PARKER
Division of Nuclear Medicine and Department of Radiology,
Beth Israel Deaconess Medical Center,
Harvard Medical School,
Boston, Massachusetts,
United States of America

12.1. PHENOMENAL INCREASE IN COMPUTING CAPABILITIES

"I think there is a world market for about five computers" — this remark is attributed to Thomas J. Watson (Chairman of the Board of International Business Machines), 1943.

12.1.1. Moore's law

In 1965, Gordon Moore, a co-founder of Intel, said that new memory chips have twice the capacity of prior chips, and that new chips are released every 18 to 24 months. This statement has become known as Moore's law. Moore's law means that memory size increases exponentially. More generally, the exponential growth of computers has applied not only to memory size, but also to many computer capabilities, and since 1965, Moore's law has remained remarkably accurate. Further, this remarkable growth in capabilities has occurred with a steady decrease in price.

Anyone who has even a little appreciation of exponential growth realizes that exponential growth cannot continue indefinitely. However, the history of computers is littered with 'experts' who have prematurely declared the end of Moore's law. The quotation at the beginning of this section indicates that future growth of computers has often been underestimated.

12.1.2. Hardware versus 'peopleware'

The exponential growth of computer capabilities has a very important implication for the management of a nuclear medicine department. The growth in productivity of the staff of a department is slow, especially when compared to the growth in capabilities of a computer. This means that whatever decision was made in the past about the balance between staff and computers is now out of date. A good heuristic is: always apply more computer capacity and less people to a new task. Or stated more simply, hardware is ‘cheap’, at least with respect to what you learned in training or what you decided last time you considered the balance between hardware and ‘peopleware’.

12.1.3. Future trends

In the near future, the increase in personal computer capability is likely to be due to an increase in the number of central processing units on a single processor chip (cores) and in the number of processing chips in a single computer. Multiple processing units have been a key feature of supercomputers for many years. Coordinating the large number of processors is often a bottleneck in the application of supercomputers for general purpose computing. Supercomputers have generally been applied to specific tasks, not to general purpose computing. The trend towards more cores in personal computers has also suffered from this bottleneck. Existing applications often run marginally faster on multicore computers than they do on a single core.

Multi-threaded programming, which has recently received more attention, ameliorates this bottleneck. Multi-threading is an efficient method of synchronizing subtasks that can be computed independently in parallel. Image processing, where different parts of the image can be processed independently, is well suited to multi-threading. Reworking just the intensive processing portions of the software as a multi-threaded application can greatly improve the overall speed on a multicore machine. In fact, several basic image processing packages are already multi-threaded. Thus, with relatively limited updating, nuclear medicine software should be able to take advantage of the trend towards multiple processors in a single computer.

An area where multiple processing units are currently used in personal computers is in graphical processing units (GPUs). GPUs, which perform the same operation on multiple parts of an image simultaneously, are classified as single instruction, multiple data processors. These units have traditionally been thought to be very difficult to program, but recently some programming tools are making them somewhat more readily accessible. They are still very difficult to program in comparison to multi-threading; however, the improved programming tools should make these processors more common for the most computing intensive data processing tasks. They are more likely to be used in the front-end computers within the imaging devices (see Chapter 11) than in workstations and servers that will be the focus of this chapter.

12.2. STORING IMAGES ON A COMPUTER

12.2.1. Number systems

12.2.1.1. Decimal, binary, hexadecimal and base 256

Everyone is familiar with the Arabic or decimal number system, which uses ten symbols, 0–9. Computer circuits are best at dealing with two values, fully on and fully off. The binary number system has two values, 0 and 1. These values can represent on/off, high/low or true/false. From right to left, the digits in a binary number represent ones, twos, fours, eights, sixteens, etc. Computers have become much better at performing conversions to decimal numbers, so that users generally do not have to worry about binary numbers; however, images are an exception. It is still useful to know something about binary numbers to understand image representation at a fundamental level.

Binary numbers are natural for computers, but they are very unnatural for humans. A compromise between humans and computers is the hexadecimal or base-16 number system. The decimal number, 1000, is represented by 3E8 in hexadecimal, which is much easier to remember than 1111101000 in binary. The hexadecimal number system uses 16 symbols: 0–9 are used for the first ten symbols and A–F are used for the last six symbols. From right to left, the digits in a hexadecimal number represent 1s, 16s, 256s, etc.

The value of a decimal digit is 10^n, where n is the position. The value of a binary digit is 2^n, and the value of a hexadecimal digit is 16^n. Four binary digits represent the same value as one hexadecimal digit: $2^4 = 16$. For this reason, one can easily convert back and forth between binary and hexadecimal, but not between binary and decimal.

Another interesting number system is used with the Internet (Table 12.1). It uses a base 256 number system, but rather than inventing a whole new set of symbols, it uses the corresponding decimal numbers as symbols. Since the decimal numbers from 0 to 255 vary in length, it needs to use periods to separate the digits. From right to left, the digits in an Internet address represent 1s, 256s, 256^2s, etc. The Internet number system is the most convenient compromise between the computer's preference for binary numbers and people's familiarity with the decimal system. An internet address can be considered to be a base 256 (2^8) number. The digits are separated by periods. Digit values are given by their decimal equivalents. One internet digit equals two hexadecimal digits or eight binary digits (see Table 12.1).

12.2.1.2. Kilo, mega, giga, tera

In the decimal system, kilo, mega, giga and tera are used to represent 1000, 1 000 000, 1 000 000 000 and 1 000 000 000 000, respectively. Using scientific notation, these numbers are 10^3, 10^6, 10^9 and 10^{12}, respectively. Twelve binary digits can be used to represent roughly the same range of numbers as three decimal digits: $2^{10} = 1024$ and $10^3 = 1000$. Therefore, these terms have been appropriated by the computer community to mean 2^{10}, 2^{20}, 2^{30} and 2^{40}.

TABLE 12.1. INTERNET ADDRESS NUMBER SYSTEM

	← 32 binary digits →							
Binary	0001	1000	0000	1001	1111	0011	0100	1110
Hexadecimal	1	8	0	9	F	3	4	E
Internet address		24.		9.		243.		78.

12.2.2. Data representation

Digital images are composed of individual picture elements, pixels, which represent a single point in the image. Each pixel is represented by a number or a series of numbers. There are several methods of representing each number.

12.2.2.1. Integer numbers

A group of binary digits (bits) can be interpreted in several different ways. The simplest is as a positive integer, numbers 0, 1, 2, 3, etc. The number of bits determines how large a number can be stored. Four bits, one hexadecimal digit, can represent 0–15. Eight bits can represent two hexadecimal digits, values 0–255. Eight bits are called a byte.

A byte is usually the smallest unit of storage that is used for a pixel. It is fairly limited, both in terms of the number of counts that can be collected in one pixel and in terms of how many colours can be specified by 1 byte. Given the architecture of modern computers, it makes more sense to use a number of bytes, e.g. 2 bytes (16 bits), 3 bytes (24 bits) and 4 bytes (32 bits). Table 12.2 shows the number of counts that can be represented using each of these formats. In nuclear medicine, a 2 byte image is sometimes called a word image. More generally, ‘word’ does not have a fixed meaning; it varies with the computer architecture. It is a less ambiguous way to describe larger integer values as 2 byte, 4 byte, etc.

A further complication is that pixels are sometimes represented by non-integral numbers of bytes. For example, computed tomography (CT) pixels are often 12 bits. Within computers, these pixels are often stored with 2 bytes, where the unused bits are set equal to zero. Some image formats even use a non-integral number of bytes per pixel.

TABLE 12.2. THE NUMBER OF COUNTS THAT CAN BE REPRESENTED IN DIFFERENT IMAGE FORMATS

	Bits	Range	Number of values
1 bit	1	0–1	2
1 byte	8	0–255	256
2 bytes	16	0–65 535	64 k
3 bytes	24	0–16 777 215	16 M
4 bytes	32	0–4 294 967 295	4 G
5 bytes	40	0–1 099 511 627 776	1 T

where

Symbol	Prefix	Power of 2	Value	Power of 10
k	kilo	2^{10}	1024	$\sim 10^{3}$
M	mega	2^{20}	1 048 567	$\sim 10^{6}$
G	giga	2^{30}	1 073 741 824	$\sim 10^{9}$
T	tera	2^{40}	1 099 511 627 776	$\sim 10^{12}$

12.2.2.2. Signed versus unsigned numbers

Positive integers are called unsigned integers. In ‘2s compliment’ arithmetic, negative integers, –1, –2, –3, etc., are represented by the largest, next largest, etc. binary numbers. The sign of the number is determined from the value of the highest order bit. If a signed number is displayed by mistake as an unsigned number, it is shown as if it had very high activity. Understanding the difference between signed and unsigned numbers makes it relatively easy to diagnose and know-how to solve this problem.

12.2.2.3. Floating point and complex numbers

Floating point representation is analogous to scientific notation where a number is represented by a mantissa times 10^n, where n is called the exponent. Instead of 10^n, computer representation uses 2^n. Using the Institute of Electrical and Electronics Engineers (IEEE) standard, both the number and the exponent are signed integers. For the single-precision, 32-bit IEEE standard, the magnitude of the largest number is about 3.4×10^{38} and the magnitude of the smallest non-zero number is about 1.2×10^{-38}. For the double-precision, 64-bit IEEE standard, the magnitude of the largest number is about 1.2×10^{308} and the magnitude of the smallest non-zero number is about 2.2×10^{-308}.

Floating point numbers are not generally used for raw nuclear medicine data; however, during processing, they can be quite useful both to represent fractions and because some processing involves large intermediary values.

Complex numbers consist of two parts, a real part and an imaginary part. They are written $x + \mathrm{i}y$ where x is the real part and $\mathrm{i}y$ is the imaginary part. The symbol i stands for $\sqrt{-1}$. In mathematics, the symbol j is often used instead of i. Each complex number is represented as two floating point numbers. Complex numbers only come up in nuclear medicine during processing. However, in magnetic resonance imaging (MRI), complex numbers are the most logical representation for the raw signals that come from the magnet.

12.2.2.4. Byte order

Memory is addressed by byte. When a number is stored using 4 bytes, then successive numbers start at addresses 0, 4, 8, etc. Confusion can arise because computer manufacturers did not adopt a standard way of assembling the bytes into a number. From least significant to most significant, some manufactures assembled 4-byte numbers using address 0 followed by address 1, e.g. 0123; other manufacturers assembled 4-byte numbers in the order 3210. The former is sometime called ‘big-endian’, the big address is at the end — the big values come first; the latter is sometimes called ‘little-endian’.

12.2.3. Images and volumes

Images can be represented by 2-D functions. If x and y are taken to be horizontal and vertical, a 2-D function, $f(x, y)$, gives the value of the image at each point in the image. Ever increasingly, imaging equipment produces not a single image, but a 3-D volume of data, e.g. right to left, anterior to posterior and caudal to cephalic. A volume of data can be represented as a 3-D function, $f(x, y, z)$.

2-D, 3-D, 4-D, etc. are often used loosely in medicine, but it can be insightful to clearly understand the dimensions involved in an application. A relevant example is human vision. A single human eye can see in only two dimensions; the retina is a 2-D structure. From parallax as well as other physiological inputs, it is possible for people with binocular vision to perceive the depth of each point in the visual image. The most appropriate model of the perceived image is a multi-valued function with value intensity, hue, saturation and depth. In this model, the function is 2-D. This 2-D function represents a surface in a 3-D volume. Over time, the mind is able to construct a 3-D model from sequential 2-D surfaces.

On a more basic level, optics models an electromagnetic signal going through an aperture as a complex 2-D function. In addition to amplitude information, which can be sensed by the eye, the function also has phase information. Holography takes advantage of phase information, allowing the observer to 'see around' objects if there is no other object 'in the way'. However, basic physics limits the amount of information to a sparse (essentially 2-D) set of data contained within the 3-D space.

One of the challenges of data visualization is to facilitate the input of 3-D data using the 2-D channel afforded by the visual system. Often, one dimension is mapped into time using cine or a mouse to sequence through a stack of images. Rendering such as a reprojection or a maximum intensity projection can also help by providing an overview or by increasing the conspicuousness of the most important features in the data. Both of these visualization methods also use a sequence of images to overcome the limitations of the 2-D visual channel.

12.2.3.1. Continuous, discrete and digital functions

The real world is usually modelled as continuous in space and time. Continuous means that space or time can be divided into infinitesimally small increments. A function $f(x)$ is said to be a continuous function if both the independent variable x and the dependent variable f are represented by continuous values. The most natural model for the distribution of a radiopharmaceutical in the body is a continuous 3-D function.

An image that is divided into pixels is an example of a discrete function. The independent variables x and y, which can only take on the values at particular pixel locations, are digital. The dependent variable, intensity, is continuous. A function with independent variable(s) that are digital and dependent variable(s) that are continuous is called a discrete function.

A computer can represent only digital functions, where both the independent and dependent variable(s) are digital. Digital values provide a good model of continuous values if the coarseness of the digital representation is small with respect to the standard deviations of the continuous values. For this reason, digital images often provide verisimilar representations of the continuous world.

Nuclear medicine is intrinsically digital in the sense that nuclear medicine imaging equipment processes scintillation events individually. The original Anger scintillation cameras produced analogue horizontal and vertical position signals, but modern scintillation cameras process the signals digitally and the output is digital position signals. Most positron emission tomography (PET) cameras have discrete crystals, so that the lines of response are intrinsically digital.

12.2.3.2. Matrix representation

The surface of the gamma camera can be visualized as being divided into tiny squares. An element in a 2-D matrix can represent each of these squares. A 3-D matrix can represent a dynamic series of images or single photon emission computed tomography (SPECT) data collection. A 3-D matrix is equivalent to a 3-D digital function, $f[x, y, z]$. Both representations are equivalent to the computer program language representation, $f[z][y][x]$.

The lines of response in a PET camera without axial collimation are more complicated. Perhaps the easiest way to understand the data is to consider a ring of discrete detectors. The ring can be represented as a 2-D array — one axial and one tangential dimension. Any line of response (LOR) can be defined by two crystals, and each of those two crystals is defined by two dimensions. Thus, each LOR is defined by four dimensions. A 4-D matrix can represent the lines of response (not all crystal pairs are in coincidence, so the matrix is sparse, but that is a detail).

It is particularly simple for a computer to represent a matrix when the number of elements in each dimension is a power of two. In this case, the x, y and z values are aligned with the bits in the memory address. Early computer image dimensions were usually powers of two. Modern programming practice has considerably lessened the benefit of this simple addressing scheme. However, where hardware implementation is a large part of the task, there is still a strong tendency to use values that are a power of two.

12.3. IMAGE PROCESSING

This section will present an introduction to the general principles of image processing.

12.3.1. Spatial frequencies

12.3.1.1. Vision and hearing

Humans conceptualize images in terms of what they perceive. Thinking of images in terms of spatial frequencies is not natural. However, the relations between perception, $f(x, y)$, and a spatial frequency representation, $F(k_x, k_y)$, can be understood by appealing to an analogy to hearing. Sound waves striking the ear-drum can be represented by a function of time, $f(t)$. A pure tone is a sinusoidal variation in pressure as a function of time. Humans do not think of a sinusoid when they hear a pure tone; it is much more natural to think of a pure tone in terms of a musical note.

The cochlea transforms sinusoidal changes in pressure over time into firing of the single nerve in the auditory nerve that represents this tone. The pressure signal, $f(t)$, is transformed to frequencies. If ω represents the frequencies, the transformed signal can be represented by the function $F(\omega)$. It is recalled that angular frequency ω is 2π times frequency. In this simple case, $F(\omega)$ is an impulse at the frequency corresponding to the tone. The process that the cochlea performs can be represented mathematically by a Fourier transform. The 'time domain' representation, $f(t)$, of the pressure signal is transformed into a 'frequency domain' signal, $F(\omega)$, carried on the auditory nerve.

Humans hear tones; they do not feel the pressure waves. Thus, the frequency domain is more natural. Since sound waves are part of every one's early education, they are less natural than tones, but do not seem foreign. Spatial frequencies do seem foreign to most people. However, the relationship between $f(t)$ and $F(\omega)$ is exactly analogous to the relationship between $f(x, y)$ and $F(k_x, k_y)$. This analogy may help physicians and other non-mathematically oriented health professionals conceptualize the relationship between space and spatial frequencies.

12.3.1.2. Spatial sinusoids

A 'pure' spatial frequency corresponds to a sinusoidal function in the spatial domain. Sinusoidal means a sine, a cosine or a combination of the two. Figure 12.1 shows a number of different sinusoidal functions.

The spatial frequency variables k_x and k_y are often given in terms of the cycles per pixel. A signal that has successive maximums and minimums in adjacent pixels will have one complete cycle in two pixels or 0.5 cycles per pixel. The spatial frequency scale is often shown from 0 to 0.5 cycles per pixel, where 0 cycles per pixel represents a constant value in the image domain and 0.5 cycles per pixel represents an image which varies from maximum to minimum in 2 pixels.

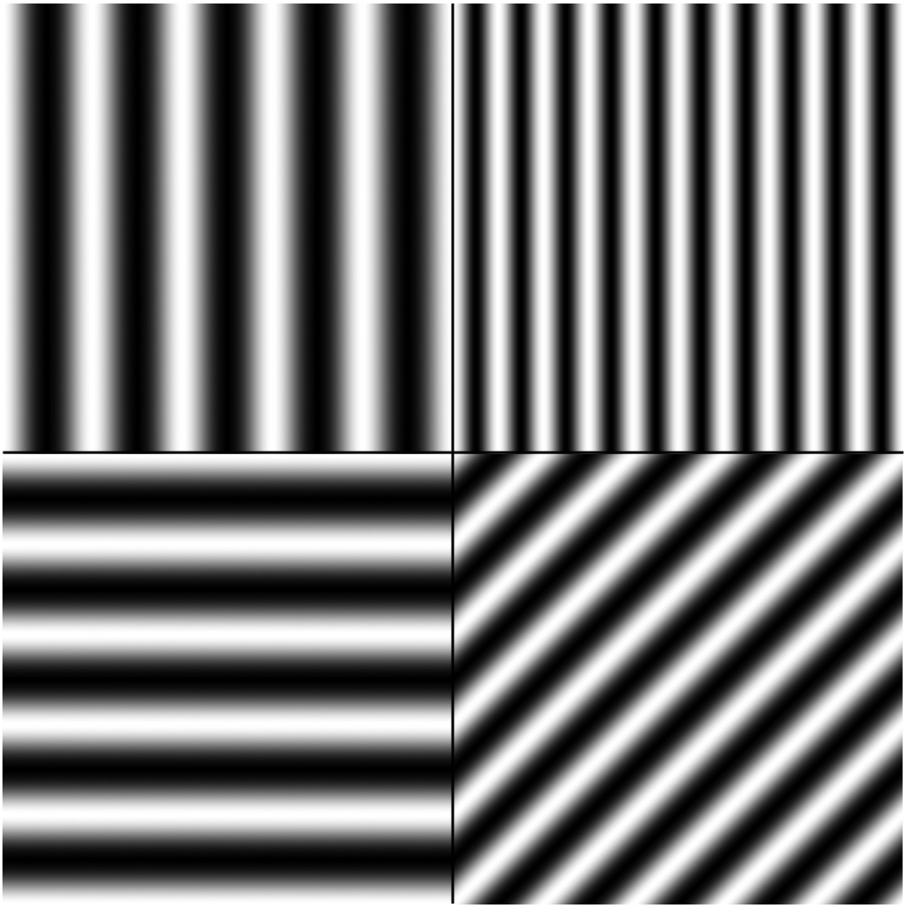

FIG. 12.1. The upper left quadrant is a sinusoid with a single frequency in the x direction; the upper right quadrant is a sinusoid with twice the frequency; the lower left quadrant is a sinusoid with a single frequency in the y direction; and the lower right quadrant is a single sinusoid which varies in both the x and y direction.

12.3.1.3. Eigenfunctions

Eigenfunctions or eigenvectors can be a difficult mathematical concept. They are often presented as an advanced concept. However, eigenfunctions are actually a basic concept. They are relatively easy to understand if the detailed mathematics is avoided.

Eigenfunctions are the ‘natural’ functions of the system. For example, an eigenfunction of a swing is a sinusoidal back and forth motion. The swing naturally wants to go back and forth in a sinusoidal motion. The output of a system when the input is an eigenfunction is a scaled version of the input. The system does not change the shape of the eigenfunction, it just scales the magnitude. For example, if a pure tone is put into a good audio amplifier, the same tone comes out, only amplified.

The model implied by these descriptions is that there is a system that has an input and an output. In the first case, the system is a swing; in the second, the system is an audio amplifier. It should be noted that eigenfunctions are properties of systems. A more relevant example is where the system is a gamma camera, the input is the distribution of radioactivity from a patient and the output is an image.

The swing and the amplifier have sinusoids as eigenfunctions. In fact, a large class of systems have sinusoids as their eigenfunctions. All linear-time-invariant or linear-shift-invariant systems have sinusoids as their eigenfunctions. Time-invariant or shift-invariant mean that the properties of the systems do not change with time or with position. Real imaging systems are rarely linear-shift-invariant systems, but there is often a region over which they can be modelled as a linear-shift-invariant system, so the sinusoids will be very useful.

A common use of eigenfunctions in nuclear medicine is in measuring the modulation transfer function (MTF). The MTF is measured by imaging a spatial sinusoid and seeing how well the peaks and valleys of the sinusoid (the modulation) are preserved (transferred by the system). If they are completely preserved, the modulation transfer is 1. If the peaks and valleys are only half of the original, the modulation transfer is 0.5. The typical bar phantom is an approximation of a sinusoidal function.

The key point for this section is that a spatial sinusoid is an eigenfunction of the imaging system, at least over some region. The image has the same shape as the input, only the amplitude (modulation) of the sinusoid is altered. The basic properties of the system can be determined by measuring the modulation transfer of the eigenfunctions. Linearity means that the effect of the imaging system on any signal that is a combination of eigenfunctions can then be determined as a scaled sum of eigenfunctions.

12.3.1.4. Basis functions

A function, $f(t)$, can be made up of a sum of a number of other functions, $g_k(t)$. For example, a function can be made from a constant, a line passing through the origin, a parabola centred at the origin, etc. This function can be written as the sum of K powers of t:

$$f(t) = \sum_{k=0}^{K-1} a_k t^k \tag{12.1}$$

These functions are just polynomials. The terms t^k in the polynomial are basis functions, with coefficients a_k. Selecting different coefficients a_k, a large number of functions can be represented.

Usually, the powers of the independent variable t^k seem like they are the most important part of a polynomial, and in many ways they are. However, it will be useful to shift the focus from the powers of t to the coefficients a_k. To make a new polynomial, the key is to select the coefficients. From this viewpoint, the t^k terms are just placeholders.

Extending this point of view, the coefficients can be thought of as a discrete function $F[k]$. The polynomial function could be rewritten:

$$f(t) = \sum_{k=0}^{K-1} F[k] t^k \tag{12.2}$$

This polynomial function provides a method of transforming from the function of coefficients to the function of time. This process is called evaluation of the polynomial. The two functions, the function of coefficients and the function of time, are equivalent representations in the sense that it is possible to transform the function of the coefficients into the function of time using this transformation. It is also possible to select K points in $f(t)$ and transform them to $F[k]$. This process, called interpolation, is beyond the scope of this chapter.

For this discussion, the key point is that the coefficient and the time representations can be converted back and forth with the processes of evaluation and interpolation. $f(t)$ can be described as a function in the 'time domain'. $F[k]$ can be described as a discrete function in the 'coefficient domain'. These two very different functions represent the same thing in the sense that using evaluation or interpolation, it is possible to convert back and forth between the representations.

12.3.1.5. Fourier domain — 'k-space'

When learning about a new way of representation of a familiar concept, it is natural to think of the familiar concept as the 'real' representation and the new representation as 'artificial'. The time or space domains are the real domains, the frequency or the spatial-frequency domains are the artificial or transform domains. However, they are really equivalent representations. One way to understand this is to think about hearing and sight. The natural domain for sight is the space domain, but the natural domain for hearing is the frequency domain. In one case, the natural domain is frequencies and the other the natural domain is just the opposite. The equivalence of the two domains is called duality. Duality is a difficult concept. Electrical engineers, who work with Fourier transforms all the time, still think of the time domain as the real domain and the frequency domain as the transform domain.

One of the best examples of duality is MRI. Chemists used to working with nuclear magnetic resonance think of the time signal as the natural signal and the frequency signal as the transform signal. Imagers think of the image from a magnetic resonance imager as the natural signal and the spatial frequencies as the transform domain. However, due to the gradients, the frequency signal gives the spatial representation and the time signal is the spatial-frequency signal. There is no ‘real’ domain and ‘transform’ domain; it all depends on the point of view.

12.3.1.6. Fourier transform

The Fourier transform equations can be written compactly using complex exponentials:

$$F(\omega) = \int f(t) e^{-i\omega t} \, dt \tag{12.3}$$

$$f(t) = \frac{1}{2\pi} \int F(\omega) e^{i\omega t} \, d\omega \tag{12.4}$$

where i is $\sqrt{-1}$.

The first equation (Eq. (12.3)) from the time or space domain to the frequency domain is called Fourier analysis; the second equation (Eq. (12.4)), going from the frequency domain to the time or space domain, is called Fourier synthesis. Fourier analysis is analogous to polynomial interpolation; Fourier synthesis is analogous to polynomial evaluation. The relation of these equations to the sine and cosine transforms can be seen by substituting for the complex exponential using Euler’s formula:

$$e^{i\omega t} = \cos(\omega t) + i \sin(\omega t) \tag{12.5}$$

These equations have been written in terms of time and frequency. Analogous equations can be written in terms of space and spatial frequency. For the 2-D case:

$$F(k_x, k_y) = \int f(x, y) e^{-i(k_x x + k_y y)} \, dx \, dy \tag{12.6}$$

$$f(x,y)=\frac{1}{2\pi}\int F(k_x,k_y)\mathrm{e}^{\mathrm{i}(k_x x+k_y y)}\ \mathrm{d}k_x\ \mathrm{d}k_y \tag{12.7}$$

The common use of the letter k with a subscript for the spatial frequency variable has led to the habit of calling the spatial frequency domain the 'k-space', especially in MRI. These equations are exactly analogous to the time and frequency equations with t replaced by x, y, and ω replaced by k_x, k_y. In the case of three dimensions, t is replaced by x, y, z, and ω by k_x, k_y, k_z.

The previous equations refer to continuous functions. The limits of integration of the integrals were not specified, but in fact, the limits are assumed to be $-\infty$ to $+\infty$. However, computer representation is digital. Computer representation is often thought of as discrete, not digital, since the accuracy of representation of numbers is often high, so that the quantification effects can be ignored (see Section 12.2.3.1). The Fourier transform equations in discrete form can be written as:

$$F[k]=\sum_{n=0}^{N-1} f[n]\mathrm{e}^{-\mathrm{i}2\pi kn/N} \tag{12.8}$$

$$f[n]=\frac{1}{N}\sum_{k=0}^{N-1} F[k]\mathrm{e}^{\mathrm{i}2\pi kn/N} \tag{12.9}$$

In image processing, the unit of n is often pixels, and the frequency unit k is often given as a fraction, cycles/pixel. The space variable n runs from 0 to $N-1$; the spatial frequency variable k runs from –0.5 cycles/pixel to (but not including) +0.5 cycles/pixel in steps of $1/N$.

12.3.1.7. Fourier transform as a correlation

Some understanding of how the Fourier transform pair works can be obtained by noting the analogy between these equations and a correlation. The correlation coefficient is written:

$$r=\frac{\sum x_i y_i}{\sqrt{(\sum x_i^2 \sum y_i^2)}} \tag{12.10}$$

where x and y are the two variables, and i indices over the number of samples. The denominator is just normalization, so that r ranges from –1 to 1. The action is in the numerator. It should be noted that the key feature of a correlation is that it is the sum of the products.

There is an analogy between the correlation equation and the Fourier transform equation. The index i is analogous to the variable t; x_i is analogous to $f(t)$; y_i is analogous to the sinusoid. The Fourier transform equation determines how much a function is 'like' the sinusoid, just as the correlation coefficient determines how much two variables are alike.

12.3.2. Sampling requirements

To represent a function, $f[n]$, with N points, N frequency values, $F[k]$, are needed; to represent a frequency domain signal, $F[k]$, with N points, N time or space values, $f[n]$, are needed. As a general heuristic, if a signal has N points in one representation, N points are needed in another representation. This heuristic assumes that the N points are independent.

A typical property of real systems is that they can only handle signal variations up to a particular maximum frequency. Thus, real signals often have a limited frequency content. The sampling theorem says that if a signal only contains frequencies up to a maximum of f, then it can be exactly reproduced from samples that are taken every $2/f$ per second. Sampling needs to be at twice the highest frequency. If the sampling is not fast enough, then the high frequencies appear to occur at lower frequencies. This process that occurs by sampling a frequency that is too low is called aliasing.

12.3.3. Convolution

Convolution can be used to describe any linear-time-invariant or linear-shift-invariant system. The input, the system and the output are each described by a function. The system function $h(t)$ is defined to be the output $g(t)$ when the input $f(t)$ is equal to a delta function $\delta(t)$.

In general, the input $f(t)$ can be conceptualized as a sequence of delta functions, $\delta(t - \tau)$, scaled by the input at time $f(\tau)$. Each delta function produces a component of the output that is the system function, $h(t - \tau)$, which has been shifted by a time equal to the time of the input and scaled by the magnitude of $f(\tau)$. As the system is time-invariant, the same response $h(t)$ can be used for an input at any time. As the system is linear, the inputs at different times can be considered separately and the separate outputs can be added. The formula for convolution is:

$$g(t) = \int h(t-\tau) f(\tau)\, \mathrm{d}\tau \qquad (12.11)$$

12.3.3.1. Shift, scale, add

Convolution can be summarized as shift, scale, add — shift the system function by an amount equal to the time of the input, scale by the size of the input and add all of the resulting shifted and scaled system functions. Convolution can be used to describe a time–activity curve from a region of interest. The arterial concentration is the input, the system function is the time–activity curve after arterial injection of a bolus of activity, and the time–activity curve after intravenous administration is the convolution of the arterial concentration and the system function.

Convolution is not limited to 1-D examples. The system function for an Anger camera is the image obtained in response to a point source of activity. A point source can be modelled as a 2-D delta function. The image of a point source will be a blob where the size of the blob is related to the camera's resolution. The blob can be shifted to every point in the field of view (FOV); it can be scaled for the amount of activity at each location; and all of the blobs can be added up. This process — shift, scale, add — will produce the final output image.

The formula for 2-D convolution is:

$$g(x,\, y) = \int h(x - x', y - y', z - z') f(x',\, y')\, \mathrm{d}x'\, \mathrm{d}y' \tag{12.12}$$

In three dimensions:

$$g(x,\, y,\, z) = \int h(x - x', y - y', z - z') f(x',\, y',\, z')\, \mathrm{d}x'\, \mathrm{d}y'\, \mathrm{d}z' \tag{12.13}$$

12.3.3.2. Mapping convolution to multiplication

In Section 12.3.1.4, on basis functions, it was noted that interpolation could be used to transform from a set of values of a signal to a representation in terms of the coefficients of a polynomial function, and that evaluation could be used to transform from the coefficients of a polynomial function into values. The coefficients can be thought of as a function in the coefficient domain and the values can be thought of as a function in the value domain. The process of multiplying two polynomials involves keeping track of coefficients of the t^0s, t^1s, t^2s, etc. The t's are essentially placeholders.

In polynomial multiplication, the coefficients of the first polynomial are shifted by the exponent of the second polynomial, scaled by the coefficients of the second polynomial, and the products are added. The process — shift, scale, add — is convolution. Multiplying polynomials is complicated. However, the process is much simpler in the value domain. Two signals are multiplied in the value domain by simply multiplying their values. In the value domain, multiplication of two signals is implemented by multiplying their values; in the coefficient domain, multiplication of two signals is implemented by convolution. Transformation of the complicated process of convolution in one domain into the simple process of multiplication in the other domain is one of the key properties of representing signals in terms of a set of basis functions.

In an exactly analogous fashion, the Fourier transform maps convolution into multiplication. The Fourier transform of the convolution of two functions is equal to the product of the Fourier transforms of the functions. Symbolically:

$$\int h(t-\tau)f(\tau)\,\mathrm{d}\tau \leftrightarrow F(\omega)H(\omega) \tag{12.14}$$

where the symbol '$\leftrightarrow$' represents Fourier transformation, the left side shows the time domain operations, and the right side shows the Fourier domain operation. The complicated integral in the time domain is transformed into a simple multiplication in the frequency domain.

At first, it may seem that the Fourier transform operation is just as complicated as convolution. However, since the fast Fourier transform algorithm provides an efficient method for calculating the Fourier transform, it turns out that, in general, the most efficient method of performing convolution is to transform the two functions, multiply their transforms and do an inverse transform of the product. Calculation of convolutions is one of the major practical applications of the Fourier transform.

12.3.4. Filtering

The word 'filtering' refers to processing data in the Fourier domain by multiplying the data by a function, the filter. Conceptually, the idea is that the frequency components of a signal are altered. The data are thought of in terms of their frequency or spatial frequency content, not in terms of their time or space content. The most efficient process in general is (i) Fourier transform, (ii) multiply by a filter and (iii) inverse Fourier transform. Convolution performs this same operation in the time or space domain but, in general, is less efficient.

If, however, a filter can be represented in the time or space domain with only a small number of components that are non-zero, then filter implementation using convolution becomes more efficient. Many image processing filtering operations are implemented by convolution. The non-zero portion of the time or space representation of these filters is called a kernel. In this section, many examples of small kernels are given, which are used in image processing.

The act of multiplying a signal by a function in the time or space domain is sometimes called windowing. Multiplying a signal by a window changes the amplitude of the time or space components of a signal. Owing to duality, filtering and windowing are essentially the same operation, and the distinction is not always maintained.

Psychologists refer to the just noticeable difference as the smallest difference that can be perceived by humans. In vision, the just noticeable difference is a function of the spatial frequency (more exactly the angular frequency). Humans have limited spatial resolution and cannot detect variations at very high spatial frequency. At somewhat lower frequency, detail can be perceived, but only if it is very high contrast. As the frequency decreases, lower and lower contrast variations can be perceived. At very low spatial frequencies, sensitivity to low contrast again decreases. The just noticeable contrast sensitivity peaks at about 2 cycles per degree and falls off rapidly for higher and more slowly for lower angular frequencies. These characteristics are important in image filtering design, which often seeks to maximize transfer of the image information to a human.

12.3.4.1. One dimensional processing

A 1-D signal $f(t)$ can be filtered with a frequency domain signal $H(\omega)$ by transforming $f(t)$ to $F(\omega)$, multiplying the signals to obtain $G(\omega) = F(\omega)H(\omega)$ and then inverse transforming $G(\omega)$ to produce the result $g(t)$. Alternately, the time domain representation of $H(\omega)$, $h(t)$, can be convolved with $f(t)$:

$$g(t) = \int h(t-\tau) f(t)\, \mathrm{d}t \tag{12.15}$$

A good example of 1-D processing is a graphic equalizer. The sliders on a graphic equalizer represent $H(\omega)$. Each slider corresponds to a range of frequencies. The tones in the input signal are multiplied by the values represented by the slider to produce the output signal.

12.3.4.2. Two- and three dimensional processing

Two dimensional processing is exactly analogous to 1-D processing with the 1-D variables t and ω replaced with the 2-D variables x, y and k_x, k_y. For 3-D processing, t is replaced by x, y, z, and ω is replaced by k_x, k_y, k_z. Four-, five-, six-, etc. dimensional processing can be performed similarly.

A property of the Fourier transform equations is separability — the transform with respect to x can be performed first followed by the transform on y. It is sometimes convenient to implement a 2-D transform by first doing the transform on the rows, and then doing the transform on the columns. All that is needed is a 1-D Fourier transform subroutine, and any number of dimensions can be implemented just by managing the indexing.

12.3.5. Band-pass filters

Band-pass filters maintain a range of frequencies while eliminating all other frequencies. An ideal band-pass filter is exactly one for some range of frequencies and exactly zero for the remaining frequencies. There is a very sharp transition between the pass- and the stop-zones. A general heuristic is that a sharp edge in one domain will tend to create ripples in the other domain. Ideal band-pass filtering often creates ripples near sharp transitions in a signal. It is, therefore, common to make the transition from the pass-zone to the stop-zone more gradual.

12.3.5.1. Low-pass, noise suppression and smoothing filters

A low-pass filter is a type of band-pass filter that passes low frequencies and stops high frequencies. In an image, the high spatial frequencies are needed for detail. Zeroing the high spatial frequencies smooths the image. By suppressing the high frequencies compared to the lower frequencies, the image noise is often suppressed. Low-pass filters both smooth an image and decrease the noise in an image.

Information theory tells us that image processing can only lower the information in an image. (An exception is when processing includes a priori information.) Noise suppression filtering lowers the information carried by the noise. It is common to say that processing ‘brings out’ a feature, but actually what it is doing is suppressing the noise which is confusing to the human viewer. All of the information about the content exists in the original image; processing merely reduces the ‘useless information’ contained in the noise.

A somewhat similar effect can often be obtained by simply moving away from the image or by making an image smaller in size. Moving away makes the high frequency noise correspond to angular frequencies where vision has low contrast sensitivity. These operations are not entirely equivalent to low-pass filtering. They may also move the important content to a point where contrast sensitivity is reduced. Instead, it is best to view an image at a size where the content is at the maximum visual sensitivity and the noise is suppressed with a low-pass filter.

A common method of implementing a low-pass filter is as a convolution with a small kernel. For example, Table 12.3 shows a 3 × 3 smoothing kernel, also commonly called a 9-point smoothing kernel. Each point in the smoothed image is made up of the scaled sum of nine surrounding points. Also shown is a 5 × 5 smoothing kernel. Each point is made up of the scaled sum of 25 surrounding points. The top of Fig. 12.2 shows the effect of the 5 × 5 kernel on a circular

TABLE 12.3. COMMONLY USED KERNELS IN IMAGE PROCESSING

Smoothing					
3 × 3		1	2	1	
		2	4	2	
		1	2	1	
Smoothing	1	2	3	2	1
5 × 5	2	4	8	4	2
	4	8	16	8	4
	2	4	8	4	2
	1	2	3	2	1
Sharpening	–1	–1	–1	–1	–1
	–1	–1	–1	–1	–1
	–1	–1	25	–1	–1
	–1	–1	–1	–1	–1
	–1	–1	–1	–1	–1
Unsharp mask	0	0	–1	0	0
	0	–1	–2	–1	0
	–1	–2	17	–2	–2
	0	–1	–2	–1	0
	0	0	–1	0	0
X gradient	0	–1	0	1	0
	0	–1	0	1	0
	0	–1	0	1	0
	0	–1	0	1	0
	0	–1	0	1	0
Y gradient	0	0	0	0	0
	–1	–1	–1	–1	–1
	0	0	0	0	0
	1	1	1	1	1
	0	0	0	0	0

region of constant intensity. The effects are relatively small and are limited to the edges of the circle. The bottom of Fig. 12.2 shows the effect on a very noisy image of the same object. In this case, the effect is much more pronounced.

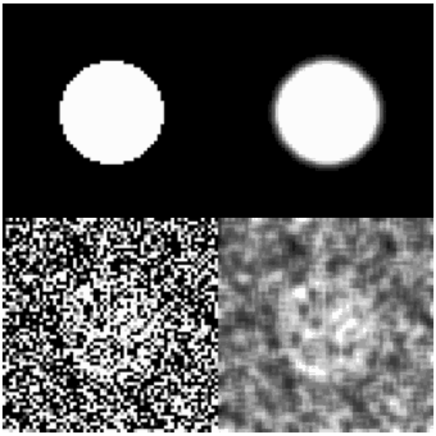

FIG. 12.2. The upper left quadrant shows a circular region of constant intensity; the upper right quadrant shows the effect of applying a 5-by-5 smoothing kernel; the lower left quadrant shows a very noisy version of the circular region; the bottom right quadrant shows the effect of applying a 5-by-5 smoothing kernel.

A common method of specifying a low-pass filter in the frequency domain is given by the Butterworth filter:

$$H(k) = \frac{1}{\sqrt{(1+(k/k_0)^{2n})}} \tag{12.16}$$

The Butterworth filter has two parameters, k_0 and n. The parameter k_0 is the cut-off frequency and n is the order. The filter reaches the value $1/\sqrt{2}$ when the spatial frequency k is equal to k_0. The parameter n determines the rapidity of the transition between the pass-zone and the stop-zone.

The filter is sometimes shown (Fig. 12.3) in terms of the square of the filter, $1/(1 + (k/k_0)^{2n})$. In that case, the filter reaches the value of half at the cut-off frequency. Confusion can arise since sometimes the filter itself is defined as the square value, i.e. without the square root. Furthermore, the filter is sometimes defined with an exponent of n instead of $2n$.

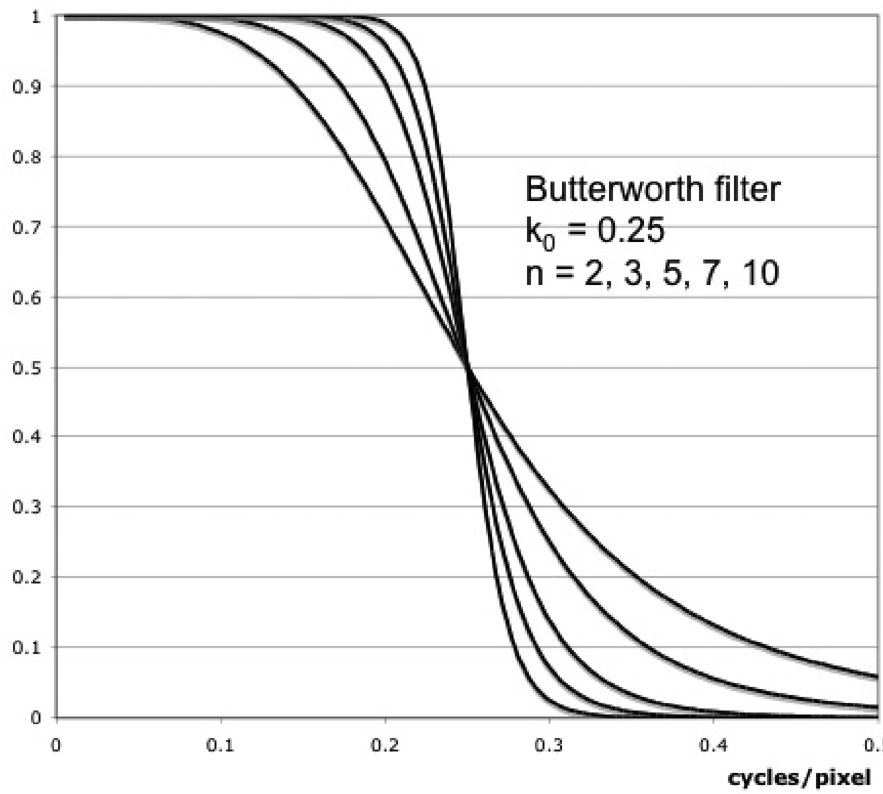

FIG. 12.3. This figure shows the square of the Butterworth filter in the spatial frequency domain with a cut-off equal to 0.25 cycles/pixel and n equal to 2, 3, 5, 7 and 10.

12.3.5.2. High-pass and edge-detection

Edges, and more generally detail, in an image are encoded with high spatial frequencies. Large regions of constant intensity are coded with low spatial frequencies. Thus, a high-pass filter will tend to emphasize the edges in an image.

A common method of implementing a high-pass filter is as a convolution with a small kernel. Table 12.3 shows a 5-by-5 sharpening kernel. The top of Fig. 12.4 shows the effect of applying this filter to a circular region with smooth edges. A high-pass filter will also emphasize pixel to pixel noise. The bottom of Fig. 12.4 shows the same circular object with a small amount of noise. Applying the high-pass filter to this image results in an image where the noise predominates. In general, a high-pass, edge-detection filter is used when the starting image has low noise.

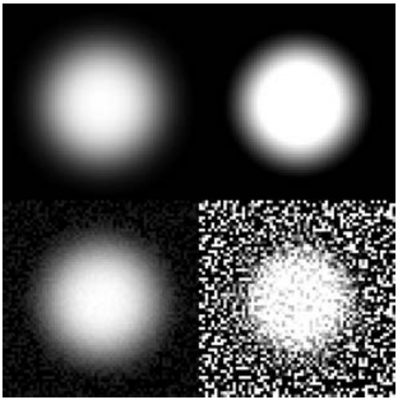

FIG. 12.4. The upper left quadrant shows a smooth circular region; the upper right quadrant shows the effect of applying the 5-by-5 sharpening kernel; the lower left quadrant shows a circular region with a small amount of noise; the bottom right quadrant shows the effect of the sharpening kernel on amplifying noise.

12.3.5.3. Unsharp masking and retinal ganglion cells

A band-pass filter, which passes intermediate frequencies and stops both low and high frequencies, will both smooth and edge enhance. It will smooth over a short range of pixels and it will enhance edges between regions at a slightly longer distance. As the low-pass and high-pass filters, this filter is often implemented in the space domain by convolution with a small kernel. Table 12.3 shows a band-pass kernel. The pixels near the centre of the kernel will be added together; those pixels will smooth the highest frequencies in the image. An annulus of pixels at a greater distance is subtracted from the centre pixels. This part of the kernel will emphasize edges that exist over this distance.

Two dimensional kernels with this form are sometimes called ‘Mexican hat kernels’ because the space of the kernel looks something like a Mexican hat. A smooth Mexican hat filter can be made from the difference of two Gaussians. This same operation is also called unsharp masking due to the similarity of a film processing method of that name.

High latitude means that images with a large dynamic range, with bright sunlight and shadows, can be displayed. Since film does not have the dynamic range of the human eye, the only way to obtain the highest latitude is to have low contrast. Even worse, a printed image, even a glossy image, has a dynamic range many orders of magnitude lower than the human eye. In an unsharp masked image, the intensity contrast over the range of the unsharp kernel is recovered from a low contrast original. The contrast between more distant objects is reduced. Although markedly altered, unsharp masked images often look verisimilar to the natural scene. In fact, they sometimes look 'better' than the natural scene. Unsharp masking is used extensively in 'coffee table books'.

In order to understand why these distorted images look verisimilar to the original scene, it is useful to consider the first ganglion cells in the retina. The first ganglion cells in the retina have a centre/surround organization. There is a small region on the retina where the cells have an excitatory input to a cell. There is a larger surrounding region where the cells have an inhibitory input to a cell. What is coded in the optic nerve is the relation of the small excitatory regions to the surrounding inhibitory regions. The ganglion cell processing is very much like unsharp masking. The brain then decodes these inputs to make a final impression of what the scene shows.

One important feature of human vision is its ability to discount illumination. To a very large extent, humans see the same colours under a wide variety of illuminations. The visual system discounts illumination. In the scene with bright sunlight and shadows, the human is aware of the illumination, but within a region, objects are still recognized even though the light reflecting from them may be orders of magnitude different depending on what part of the scene they are in. It is likely that this centre/surround mechanism is key for discounting illumination.

The important thing to remember in terms of image processing is that even though an unsharp masked image may appear to humans as verisimilar to the natural scene, it has been considerably altered from the original scene.

12.3.6. Deconvolution

Deconvolution is the process of undoing convolution. If we know the input, $f(t)$, to a system, $h(t)$, and the output, $g(t)$, from the system, then deconvolution is the process of finding $h(t)$ given $f(t)$ and $g(t)$. In the frequency or spatial frequency domain, deconvolution is straightforward — a simple division:

$$H(\omega) = \frac{G(\omega)}{F(\omega)} \tag{12.17}$$

There is a problem with this simple deconvolution. If $F(\omega)$ is zero for any ω, then $H(\omega)$ is infinite. Furthermore, since nuclear medicine deals with real systems that have noise, any time $F(\omega)$ is small, then any noise in $G(\omega)$ is amplified. The solution is to filter the simple deconvolution by a filter that depends on the power in $F(\omega)$.

Since signals usually have less power at high frequency or high spatial frequency, the first practical deconvolution was performed by simple high-pass filtering. A Metz filter (Section 12.3.7.2) can be used to emphasize the mid-frequency range while attenuating the high frequencies. If the statistical properties of the signal and noise are known or can reasonably be approximated, then a Wiener filter (Section 12.3.7.3) can be used.

12.3.7. Image restoration filters

Some filters attempt to restore degraded images. Degradations can come from many sources, but the most important for this book is degradations caused by imaging instruments such as gamma cameras. If the imaging system is linear-shift-invariant, then it can be modelled by convolution. A restoration filter needs to undo the effect of the imaging system, deconvolving for the system function. Even in the case where the system is not linear or shift-invariant, it is often possible to ameliorate the effects of a system with a filter.

12.3.7.1. Ramp

The effect of a projection/back projection is to smooth an object (see Section 12.3.5.1). In polar coordinates, the system function is given by the simple equation:

$$H(k,\theta)=\frac{1}{k} \tag{12.18}$$

It should be noted that the system function is not a function of θ. The higher frequencies in the image are reduced by a factor proportional to their spatial frequencies.

The obvious method of restoring the original image is to multiply it by a restoration filter given by:

$$G(k,\theta)=k \tag{12.19}$$

The restoration filter increases in magnitude in proportion to the spatial frequency. Owing to the shape of the filter, it is called a ramp filter. A ramp filter reverses the projection/back projection operation.

12.3.7.2. Metz

If the noise in the detected signal is uniform, then a restoration filter will alter the noise spectrum. For small restorations, the signal to noise may still be favourable, but for frequencies with large restorations, the noise may dominate. The best result is often to restore small effects completely, but restore large effects less. The Metz filter combines these two goals into a single filter. For any system function, $H(k, \theta)$, the Metz filter is given by:

$$G(k, \theta) = H^{-1}(k, \theta)(1-(1-H(k, \theta)^2)^x) \tag{12.20}$$

The system function is often written as MTF(k, θ) to emphasize that it is an MTF.

The first term in this equation, $1/H(k, \theta)$, reverses the effect of the system. When $H(k, \theta)$ is nearly one, the second term is about equal to one; when $H(k, \theta)$ is nearly zero, the second term is about equal to zero; at intermediate values, the second term transitions smoothly between these two values. In Fig. 12.5, a simulated system function is shown by the dotted lines. Four Metz filters with different X parameters are shown. At low frequencies, the Metz filter counteracts the effects of the imaging system. At high frequencies, it takes on the character of a low-pass filter. The transition between characters is controlled by the parameter X.

12.3.7.3. Wiener

Wiener filtering is used when the statistical properties of the signal and of the noise are known. A function known as the power spectral density, $S(\omega)$, gives the expected amount of power in a signal as a function of frequency. 'Power' means that the function is related to the square of the magnitude of a signal. The Wiener filter is given by:

$$H(\omega) = \frac{S_f(\omega)}{S_f(\omega) + S_n(\omega)} \tag{12.21}$$

where $S_f(\omega)$ is the power spectral density of the signal $f(t)$, and $S_n(\omega)$ is the power spectral density of the noise $n(t)$. Under the proper circumstances, it can be shown that the Wiener filter is the optimal estimate of a process $f(t)$ given the noisy data, $f(t) + n(t)$.

For those frequencies where the signal is much larger than the noise, the Wiener filter is equal to one. For those frequencies where the noise is much larger than the signal, the Wiener filter is equal to zero. The Wiener filter transitions smoothly from the pass-zone to the stop-zone when there is more noise in the data than signal. The Wiener filter provides a solid mathematical basis for the heuristic that has already been used several times.

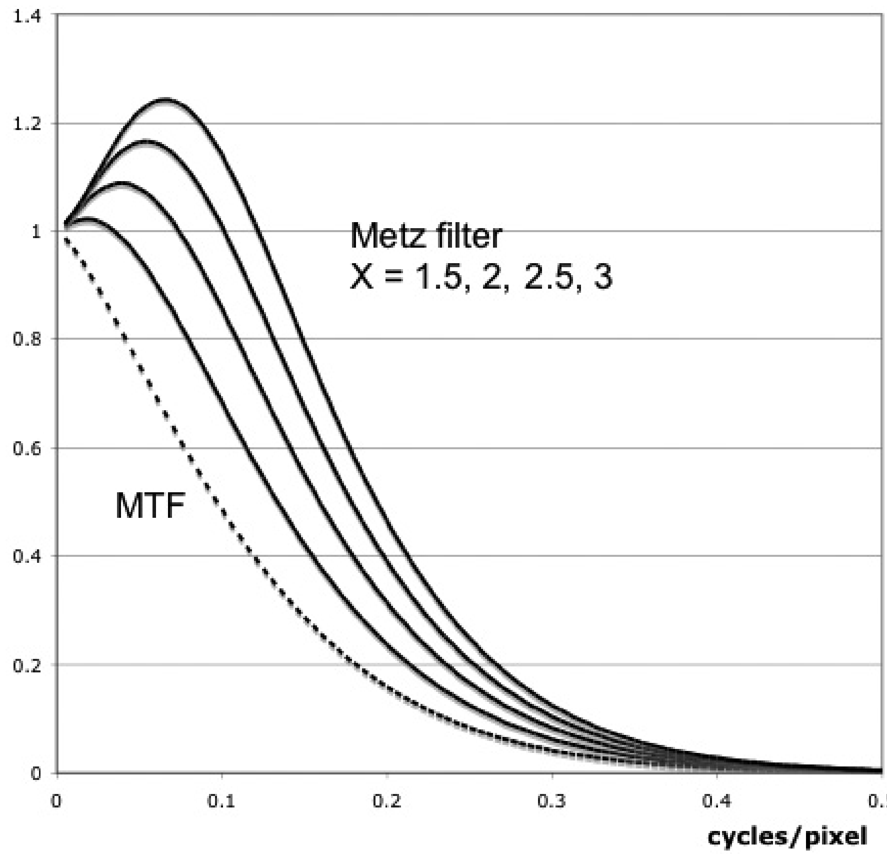

FIG. 12.5. The dotted line shows a simulated modulation transfer function (MTF). The four solid lines show Metz filters for X equal to 1.5, 2, 2.5 and 3.

12.3.8. Other processing

This chapter is predominantly concerned with linear-time-invariant or linear-shift-invariant signals. Fourier transform methods can be applied to these signals. Fourier transforms provide considerable insight and using the fast Fourier transform algorithm, they are very efficient computationally. However, the types of system that can be modelled are limited. Any real imaging system typically has an FOV. At the edge of the FOV, the system is no longer shift-invariant. Furthermore, attenuation, scatter and depth dependent resolution cannot be modelled using linear-shift-invariant systems.

12.3.8.1. Linear systems

Linear systems without being limited to time- or shift-invariance are much more powerful. Limited FOV, attenuation, scatter and depth dependant resolution can all be modelled. In addition, non-standard imaging geometries such as limited angle tomography can be modelled.

The relation between the input and the output of a linear system is equivalent to a system of simultaneous equations. An efficient method of writing a set of simultaneous equations is to use matrix algebra:

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} \tag{12.22}$$

where $\boldsymbol{f}$ and $\boldsymbol{g}$ are vectors and $\boldsymbol{H}$ is a matrix.

In the case of an image or especially a volume, the vectors $\boldsymbol{f}$ and $\boldsymbol{g}$ represent the whole image or volume. In the case of a 256 × 256 × 256 volume, each vector has 16 mega elements. The matrix, which is 16 mega elements by 16 mega elements, has 256 tera elements. It should be obvious that this simple expression hides a great deal of complexity. However, it does bring out the analogy between linear systems and linear-time-invariant or linear-shift-invariant systems.

12.3.8.2. Non-linear systems

The output of a linear system to the sum of two signals is equal to the sum of the outputs from each signal separately. Furthermore, the output of a system to a signal multiplied by a number is equal to the output due to the signal multiplied by the same number. Real systems are almost never linear. There is usually some size input for which real systems break down. Above that magnitude, increases in the input are not reflected in the output.

Non-linear systems are very difficult to deal with. Consequently, non-linear systems are often modelled by restricting their application to the range where the system is nearly linear. Other times, it is possible to transform the input or the output in order to make the system behave approximately as a linear system. There are methods of dealing with non-linear systems, but since they are usually very difficult, non-linear systems are rarely modelled directly.

12.4. DATA ACQUISITION

Modern nuclear medicine data acquisition devices have considerable computing power embedded in them that performs many tasks to improve images. These interesting and important computing functions are described in Chapter 11. This chapter will treat the camera as a black box and begin with the data that have already been obtained.

12.4.1. Acquisition matrix size and spatial resolution

For planar imaging, the FOV of a camera is divided into discrete elements that correspond to the pixels in a digital image. Most commonly, a rectilinear system is used and the digital image is analogous to a matrix (see Section 12.2.3.2). The size of the FOV and the spatial resolution of the camera are important in choosing the size of the matrix to be used. If the imaging device is modelled as a linear space-invariant system, then to avoid aliasing, the sampling should be twice the highest frequency that can be recorded by a camera (see Section 12.3.2). For example, if the FOV of the camera is 40 cm and the camera can detect frequencies up to 2 cycles/cm, then the matrix should have:

$$2 \text{ samples/cycle} \times 2 \text{ cycles/cm} \times 40 \text{ cm} = 160 \text{ samples}$$

For a normal large FOV Anger camera, a 256 × 256 matrix is more than sufficient for most imaging situations. For clinical imaging, a 128 × 128 matrix is often sufficient. For rapid dynamic images that are often very count limited, a 64 × 64 matrix can be used. For vertex to thigh imaging, a 1:4 matrix, e.g. 256 × 1024, will cover 160 cm. When including the legs, a non-power-of-two matrix may be most logical.

In the past, when memory and storage were more expensive, there was emphasis on using as small a matrix as possible. Currently, it makes more sense to emphasize that the use of a matrix is large enough to be sure that imaging is not limited.

12.4.2. Static and dynamic planar acquisition

Many of the processes of interest in nuclear medicine involve a changing pattern of distribution of a radiopharmaceutical in the body. In matrix-mode, a sequence of images collected over time allows visualization and measurement of the biodistribution of the radiopharmaceutical. Each image is often called a frame of data. Collection of a sequence of images over time is called a dynamic acquisition; collection of an image at a single point in time is called a static acquisition. Several static images may be grouped together as a multiple static acquisition; however, the difference between multiple static and dynamic acquisitions is largely conceptual. A dynamic acquisition that shows vascular delivery of the radiopharmaceutical is often called a flow study. A dynamic collection during renal scintigraphy is called a renogram.

The FOV of the gamma camera is divided into discrete rectangles that are generally square. Each rectangle corresponds to one pixel in the image and each is represented by one memory location. The size of the rectangles puts a limit on the resolution that is achievable, but the rectangle size is not the resolution. The resolution of a gamma camera is the overall ability to resolve adjacent sources of activity. For each scintillation, the gamma camera identifies the rectangle corresponding to the location of the scintillation and increments the corresponding memory location.

12.4.3. SPECT

SPECT is typically performed by acquiring a sequence of 2-D projection images from multiple different angles around the object. The 2-D projections can be represented by $p[x', z]$, where z is the axial position. The set of projections can be described by a 3-D function, $p(x', z, \theta)$, where θ is the position of the camera head.

12.4.3.1. Sinogram

The first step in reconstruction is to reformat the data in terms of axial position. The reformatted data can be described by a function $p'(x', \theta, z)$. For each axial position, the set of data, $p'(x', \theta)$, has one line of data from each of the projections. Each of the images, $p'(x', \theta)$, is called a sinogram. The x position in the sinogram is the same x position in the raw projection image. The other position θ is the projection angle. A point source will project onto all of the projection images at the same axial position z. The position of the point source in the θ direction will form a sine wave in the sinogram, hence the name sinogram.

The sinogram is not a particularly useful way of visualizing the data, but it is frequently available in nuclear medicine systems. It allows easy identification of a few artefacts. If there is patient movement during data acquisition, there will be discontinuity of the sine waves at the projection angle θ where the movement occurred. A detector problem will show up as straight lines of different intensity in the θ direction. Intermittent problems with the whole detector will result in straight lines in the x direction.

12.4.3.2. Sampling requirements in SPECT

Each of the sinograms $p'(x', \theta)$ needs to be reconstructed into an image $f(x, y)$ where x and y are coordinates with respect to the patient. The two types of reconstruction are Fourier reconstruction and iterative reconstruction. For an adequately sampled linear-shift-invariant system, Fourier reconstruction is exact in the absence of noise. Iterative reconstruction is approximate, but allows a much more realistic model of the system including attenuation, scatter, depth dependent collimator resolution and non-regular sampling.

A good heuristic is that in order to reconstruct N points in a slice, N data samples are needed. Sampling tangentially to the slice is reasonably straightforward; x' should have the same sampling rate as x and y. A good rule for Fourier reconstruction is that the number of angular samples over 180° is $\pi n/2$, where n is the samples along x'. This number is equal to the heuristic within a factor of $\pi/2$.

Resolution recovery is used in several nuclear medicine applications. Resolution recovery improves 'resolution' compared to the heuristic given above. The extra information required to improve 'resolution' is supplied by an a priori assumption, e.g. the edges in the image are relatively sharp. If the object does not meet the a priori assumption, then artefacts will be produced. As long as the assumptions are reasonable, resolution recovery can be useful.

Well designed collimators always involve a trade-off between resolution and sensitivity. A good heuristic is that for optimal detection, the collimator should have a resolution equal to the object to be detected. However, there may be more than one resolution object to be detected. For example, in cardiac imaging, if a low resolution initial image is needed for positioning followed by a high resolution image, then using resolution recovery for the higher resolution image can ameliorate use of a lower resolution collimator.

Each tomographic data sample includes a combination of the voxels in the object. For detectors that measure energy, so-called square-law detectors, the data samples will always be sums of object voxels. (This limitation is not true for MRI where the data have both an amplitude and a phase.) For tomographic data, square-law detectors always oversample low frequencies and reconstruction must amplify the high frequencies compared to the low frequencies. Typically, noise is relatively uniform in the projection space. Thus, after reconstruction, the high frequency noise is amplified with respect to the low frequency noise.

12.4.4. PET acquisition

The detectors in positron cameras detect single photons. However, the key feature of a positron camera is detection of pairs of photons in coincidence. Two photons detected in coincidence define an LOR. The LORs are then reconstructed into a 3-D image set. At that point, PET and SPECT data are the same.

Most PET cameras are made from individual crystals. The 'singles' data from the crystals can be recorded in a number of memory elements equal to the number N of crystals. $N \times N$ memory elements are needed for the LORs. (Some potential LORs do not traverse the imaging area.) The crystals in a PET camera make a 2-D array. In the usual cylindrical geometry, the two dimensions are the axial slice and the angular position around the ring of detectors for that slice. LORs are defined by two crystals and four dimensions. To limit the electronics, early PET cameras limited coincidences to a single slice. In that case, the slice position is the same for both detectors, and the LORs are 3-D, one axial and two angular position dimensions.

12.4.4.1. '2-D' and '3-D' volume acquisition

Modern PET cameras can collect oblique LORs, greatly increasing sensitivity. Some cameras have axial collimators that can be inserted or retracted while others do not. Imaging with axial collimation involves a 3-D object, a 3-D image and 3-D data. Strangely, this type of imaging is called '2-D'. Imaging without axial collimation again involves a 3-D object and a 3-D image, but in this case the data are 4-D. This type of imaging is called '3-D' imaging. Although these terms for imaging with and without axial collimation do not make sense, they are imbedded in the PET vernacular.

12.4.4.2. Time of flight

Time of flight (TOF) imaging measures the difference in arrival of the two photons at the two detectors. The difference in arrival time can be used to position the annihilation event along the LOR. This provides an additional dimension to the data. Non-axial collimation, TOF PET data are 5-D. The resolution along the LOR is relatively low. The speed of light is about 3×10^8 m/s, which means that in 1 ns photons travel 30 cm. Current detectors can detect differences in arrival of about 0.5 ns. Thus, improvements using TOF detection are important, but modest.

12.4.4.3. Sampling requirements in PET

Data sampling is determined by the LORs. Determination of the resolution of a PET camera is complicated, but once it is determined, then the size of the 3-D reconstructed matrix is the same as in SPECT imaging. Non-axial collimation PET adds oblique data with the potential of increasing the axial dimension. In practice, the axial resolution is similar for axial collimation and non-axial collimation PET. With cylindrical cameras, there are many more oblique angles that can be detected on the centre slice than on end slices. In non-axial collimated PET, the end slices are noisy. The non-uniform axial sampling is ameliorated in whole body imaging by overlapping the different bed positions, so that the data are more uniformly sampled.

12.4.5. Gated acquisition

Data acquisition can be gated to a physiological signal. Owing to count limitation, if the signal is repetitive, the data from multiple cycles are usually summed together. A gated, static acquisition will have three dimensions — two spatial and one physiological. A gated dynamic acquisition will have four dimensions — two spatial, one time and one physiological. A gated SPECT or PET acquisition will have a different four dimensions — three spatial and one physiological. A gated dynamic SPECT or PET acquisition will have five dimensions — three spatial, one time and one physiological. Count limitations will often limit the dimension and/or the elements per dimension.

12.4.5.1. Cardiac-synchronized

Gated cardiac studies were one of the early studies in nuclear medicine. Usually, the electrocardiogram is used to identify cardiac timing. The R-wave is relatively easy to detect; however, at least a little knowledge of electrocardiography is useful to know how to select electrode positions where the R-wave has a high amplitude.

The timing of events during the cardiac cycle changes with cycle length. The timing of contraction, systole, and the initial part of relaxation, diastole, change relatively little as cycle length changes. Most of the change in cycle length is a change in diastasis, the period between early relaxation and atrial contraction. Atrial contraction, atrial systole, has a relatively constant relation to the end of the cardiac cycle, i.e. to the next R-wave. Thus, for evaluation of systolic and early diastolic events, it is better to sum cycles using constant timing from the R-wave than to divide different cycles using a constant proportion of the cycle length.

However, with constant timing, different amounts of data are collected in later frames depending on the number of cycles that are long enough to reach those frames. Normalizing the data for the acquisition time for each frame can ameliorate this problem. Alternatively, systole and early diastole can be gated forwards in time from the preceding R-wave, and atrial systole can be gated backwards in time, joining the two during diastasis.

12.4.5.2. Respiratory-synchronized

Gating to respiration has been used in nuclear medicine both to study respiratory processes and to ameliorate the blurring caused by breathing. The signal to use for respiratory synchronization is not clear-cut. Successful synchronization has been based on different types of plethysmography, on chest or abdominal position changes, and on image data. Plethysmography, measurement of the air in the lungs, can be measured by integrating the rate of airflow or indirectly from changes in temperature of the air as it is breathed in and out. Corrections often need to be made for errors that build up from cycle to cycle. Electrical impedance across the chest changes with respiration and can also be used to detect lung volume changes. Chest and abdominal motion often correlate with the respiratory cycle, although detection of motion and changes in diaphragmatic versus chest wall breathing also pose problems. Although difficult to measure, respiratory gating is becoming more common, particularly in quantitative and high resolution PET applications.

12.4.6. List-mode

Instead of collecting data as a sequence of matrixes or frames, the position of each scintillation can be recorded sequentially in a list. Periodically, a special mark is entered into the list that gives the time and may include physiological data such as the electrocardiogram or respiratory phase.

List-mode data are generally much less efficient than matrix-mode data. For example, a 1 million-count frame of data collected into a 256 by 256 matrix will require 64 000 memory locations. In list-mode, the same data will require 1 million memory locations. The reason list-mode is so much less efficient is that it contains extra information, namely the order of arrival of each of the scintillations. However, this information is not interesting. List-mode is more efficient than matrix-mode when there is less than one event per pixel on average.

Since an image with one event per pixel is very noisy, it may be surprising that list-mode is useful at all. However, with gated images, multiple individual images are added to produce one output image. List-mode can be more efficient if it is necessary to keep the separate images that make up one gated image for post-processing.

12.5. FILE FORMAT

"The nicest thing about standards is that there are so many of them to choose from." — Ken Olsen, founder of Digital Equipment Corp.

12.5.1. File format design

12.5.1.1. Interfaces

Standards have facilitated the rapid development of computers. Standards can be considered as part of a larger topic, interfaces. In general, an interface is a connection between an entity and the rest of the system. If each entity in the system has an interface and if the interface is the only allowed way for the other parts of the system to interact with an entity, then there is a great simplification of the overall system. Development of one part of a system depends only on that part and the interface. It does not depend on any other part of the system. Each part is greatly simplified because it is separate from the rest of the system. A system that is divided up into parts with well designed interfaces between the parts is called modular.

There was an early problem with development of complex systems. As more resources such as programmers were added to a task, very little additional output resulted. The problem was that as the task grew in complexity, more time was spent on communication between the programmers and less was spent on programming. At times, the output actually decreased as more resources were devoted to the task. The problem is that as a project becomes larger, the complexity tends to increase exponentially.

In a modular system, the details of each portion of the task are hidden from other parts of the task. Data hiding is a goal for good system design. Modularizing a task tends to linearize the complexity. As a task with complexity n becomes twice as large, the work becomes about $2n$, not n^2 or e^{2n} times as much.

File formats are a type of interface. They define how information will be transferred from one system to another. By having a well designed format, each system becomes separate. The goal of this section is to describe existing file formats and understand how they can simplify the design of a nuclear medicine information system.

12.5.1.2. Raster graphics

There are two general image formats — raster graphics and vector graphics. Vector graphics define an image in terms of components such as points, lines, curves and regions, e.g. polygons. Points, lines, curves and the boundaries of regions can have a thickness and a colour. The interiors of regions can have colours, gradients or patterns. For images that are made up of these types of component, the vector graphics description is very compact. Vector graphics can be smoothly expanded to any size. Modern type faces, which are defined in terms of vector graphics, can be expanded to any size while retaining their smooth outlines. Vector graphics are used extensively in gaming software. Vector graphics are not useful for nuclear medicine images, so this chapter will only describe raster graphics.

Raster graphics images are made up of a matrix of picture elements or pixels. Each pixel represents one small rectangle in the image. The primary data in nuclear medicine are numbers of counts. Counts are translated into a shade of grey or a colour for display. There are many ways the shade of grey or the colour of a pixel can be encoded. One factor is the number of bits that are used for each pixel. Grey scale images often have 8 bits or 1 byte per pixel. That allows 256 shades of grey, with values from 0 to 255.

A common way to encode colour images is in terms of their red, green and blue (RGB) components. If each of the colours is encoded with 8 bits, 24 bits or 3 bytes are needed per pixel. If each colour is encoded with 10 bits, 30 bits are needed per pixel. RGB encoding is typical in nuclear medicine, but there are other common encodings such as intensity, hue and saturation. For printing, images need to be encoded in terms of cyan, magenta and yellow. Use of more complicated encodings allows more vibrant images that use additional colours, e.g. black, orange, green, silver or gold.

12.5.1.2.1. Transparency

Graphic user interfaces often allow a composition of images where background images can be seen through a foreground image. In such cases, each pixel in an image may be given a transparency value that determines how much of the background will come through the foreground. A common format is 8 bits for each of the RGB colours and an 8-bit transparency value, giving a total of 32 bits per pixel.

12.5.1.2.2. Indexed colour

Typically, an image will include only a small number of the 16 million (2^{24}) possible colours in a 24 bit RGB palette. A common method of taking advantage of this is to use indexed colour. Instead of storing the colour in each pixel, an index value is stored. Typically, the index is 8 bits, allowing 256 colours to be specified. In the case where more than 256 colours are used, it is often possible to approximate the colour spectrum using a subset of colours. Sometimes, a combination of colours in adjacent pixels will help to approximate a broader spectrum. The algorithms for converting full spectrum RGB images into a limited palette are remarkably good. It often requires very careful inspection of a zoomed portion of the image to identify any difference.

For indexed colour, a colour palette is stored with the images. The colour palette has 256 colour values corresponding to the colours in the image. The 8 bit index value stored in a pixel is used to locate the actual 24 bit colour in the colour palette, and that colour is displayed in the pixel. Each pixel requires only 8 bits to store the index as opposed to 24 bits to store the actual colour. The colour palette introduces an overhead of 8 bits for each of the 256 colours, but since most images have tens of thousands or hundreds of thousands of pixels this overhead is small.

12.5.1.2.3. Compression

The information content of an image is often much smaller than the information capacity of an image format. For example, images often have large blank areas or areas that all have the same colour value. Therefore, it is possible to encode the image using less space. Improving efficiency for one class of images results in decreasing efficiency for another type of image. The trick is to pick an encoding which is a good match for the range of images that are typical of a particular application.

One of the simplest and easiest encodings to understand is run-length encoding. This method works well for images where large areas all have the same value, e.g. logos. Instead of listing each pixel value, a value and the number of times it is repeated are listed. If there are 50 yellow pixels in a row, rather than listing the value for yellow 50 times, the value ‘50’ is followed by the value for yellow. Two values instead of 50 values need to be listed.

Another common encoding is called Lempel–Ziv–Welch (LZW) after its developers. It was originally under patent, but the patent expired on 20 June 2003, and it may now be used freely. The LZW algorithm is relatively simple but performs well on a surprisingly large class of data. It works well on most images and works particularly well on images with low information content, logos and line drawings. It is used in several of the file formats described below.

Both run-length encoding and LZW encoding are non-destructive or reversible or lossless coding; the original image can be exactly reconstructed from the encoded image. After destructive or irreversible or lossy coding, the original image cannot be exactly recovered from the coded image. However, much more efficient coding can be performed with destructive encoding. When non-destructive encoding results in reduction of the image size by a factor of 2–3, destructive encoding will often result in a reduction of the image size by a factor of 15–25. The trick is to pick an encoding system where the artefacts introduced by the coding are relatively minor.

The human visual system has decreased sensitivity to low contrast, very high resolution variations. Details, high resolution variations, are conspicuous only at high contrast. Discrete cosine transform (DCT) encoding takes advantage of this property of the visual system. DCT encoding is in the Joint Photographic Experts Group (JPEG) standard. The low spatial resolution content of the image is encoded with high fidelity, and the high spatial resolution content is encoded with low fidelity. The artefacts introduced tend to be low contrast detail. High contrast detail, and both low and high contrast features at low resolution are faithfully reproduced. Although there are artefacts introduced by the coding, they are relatively inconspicuous for natural scenes. For nuclear medicine images, the artefacts are more apparent on blown up images of text.

Wavelets are a generalized form of sinusoids. Some improvement in compression ratios at the same level of apparent noise can be obtained using wavelet-transform coding. The ‘blocky’ artefacts that degrade DCT images are not seen with wavelet-transform images. The JPEG 2000 standard uses wavelet-transform coding.

Non-destructive coding systems often make use of the fact that adjacent pixels are equal. The statistical noise in nuclear medicine images reduces the similarity of adjacent pixels, thus reducing the utility of non-destructive coding. Destructive coding may overcome this limitation, and since the statistical variations generally do not carry any significant information, image quality may not be greatly degraded.

12.5.2. Common image file formats

Common image file formats could be used for some or all of the raw nuclear medicine image data. In fact, they are rarely used for this purpose. However, secondary images, especially when used for distribution of image information, generally use these standard file formats. This section will sketch the format of these files, the advantages and disadvantages of the formats, and the typical uses of these formats.

12.5.2.1. Bitmap

Bitmap (BMP) is a general term that may refer to a number of different types of data. When used in reference to image formats, it may be used to mean a raster graphic as opposed to a vector graphic format, but it often refers to a Windows image format. The Windows file format is actually called a device independent bitmap (DIB). The external DIB file format is distinguished from various device dependent internal Windows BMPs. File name extensions .bmp and .dib are used for the BMP image file format. BMP may be used to imply an uncompressed format, but the DIB format defines several types of compression.

12.5.2.2. Tagged Image File Format

Aldus Corp. created the Tagged Image File Format (TIFF). Adobe Systems Inc., which merged with Aldus, now owns the copyright. The TIFF format can be used free of licensing fees. The TIFF format includes many types of image. Most commonly, it is used for 8 bit grey scale, 8 bit indexed colour or 24 bit RBG colour images. Images are typically compressed with the non-destructive LZW algorithm.

The TIFF format is often used for high quality single images that are non-destructively compressed. Although multiple other uses are possible, this application is often thought of as the strength of the TIFF format. The wide variety of options provided by the TIFF format is both a strength and a weakness. Few programs support all of these options, so that a TIFF image produced by one program may not be readable by another program. A particular weakness for nuclear medicine data is that the multiframe option is rarely supported.

12.5.2.3. Graphics Interchange Format

CompuServe, now a subsidiary of AOL, developed Graphics Interchange Format (GIF). GIF is a relatively simple indexed format with up to 8 bits per element. Up to 256 colours are selected from a 24 bit RGB palette. Pixels can be transparent, in which case the background colour is displayed. The 89a specification included multiple images that may be displayed as a cine. Compression is performed with the LZW algorithm.

As the GIF format is relatively simple yet quite useful, it is very commonly supported. All web browsers support this format. It is particularly good for storing low information content pictures such as logos and clip art. It is also good for images such as nuclear medicine images. Cine capability is widely supported, so it is very convenient for showing dynamic, gated and 3-D data.

12.5.2.4. Joint Photographic Experts Group

The Joint Photographic Experts Group developed the JPEG format in 1992. JPEG was approved in 1994 by the International Organization for Standardization as ISO 10918-1. The JPEG standard leaves some issues unspecified. The JPEG File Interchange Format (JFIF) clarifies these issues. To indicate that a JPEG image also follows this standard, it is sometimes called a JPEG/JFIF image, but generally JFIF is assumed. The JPEG standard is a reasonably simple, non-indexed grey scale and colour image format that allows adjustable destructive compression.

The JPEG coding tries to match the coding to human vision. It uses more precision for brightness than for hue. It uses more precision for low frequency data than for high frequency detail. The JPEG format is particularly good at compressing images of natural scenes, the type of images in routine photography. All web browsers support this format, and it is very widely used in general photography products. It is not a particularly good format for line drawings and logos; the GIF format is better for these types of image.

12.5.3. Movie formats

Multitrack movie formats allow audio and video to be blended together. Often, there are many audio tracks and many video tracks where the actual movie is a blending of these sources of information. A key part of a multitrack format is a timing track, which has information about how the tracks are sequenced and blended in a final presentation. However, in nuclear medicine, there is rarely a need for all of this capacity. Often, all that is needed is a simple sequence of images. A more complex movie format can be used for this type of data, but often a simpler format is easier to implement.

12.5.3.1. Image sequence formats

Of the image formats described so far, BMP, TIFF and GIF can be used to define an image sequence. An extension of the JPEG format, JPEG 2000, also allows image sequences. As with the multiframe version of BMP and TIFF formats, JPEG 2000 is relatively poorly supported. However, the multiframe version, 89a, of the GIF format is widely supported. It is supported by all web browsers and by almost all image processing and display programs. The GIF format is a logical choice for distribution of cine images.

12.5.3.2. Multitrack formats

There are several multitrack movie formats available. As newer formats that are still under development, they tend to be controlled by particular vendors. The AVI format is controlled by Microsoft, the QuickTime format is controlled by Apple, the RealVideo format is controlled by RealNetworks and the Flash format is controlled by Adobe.

12.5.4. Nuclear medicine data requirements

There are two types of information in a nuclear medicine study — image information and non-image information. As described in the previous section, there are several general purpose image and movie formats. Often, these formats lack capabilities that would be optimal for medical imaging. However, some formats, e.g. TIFF, are general enough to be used for the image portion of the study data. The advantage of using a widely accepted format is that it allows the great diversity of software available for general imaging to be adapted to medical imaging.

12.5.4.1. Non-image information

There is unique nuclear medicine information that must be carried along reliably with the images. This information includes: identification data, e.g. name, medical record number (MRN); study data, e.g. type of study, pharmaceutical; how the image was acquired, e.g. study date, view; etc. This information is sometimes called meta-information. Most general image formats are not flexible enough to carry this information along with the image.

12.5.4.1.1. American Standard Code for Information Interchange

Text information is usually and most efficiently coded in terms of character codes. Each character, including punctuation and spacing, is coded as a number of bits. Initially, 7 bits were used; 7 bits allowed $2^7 = 128$ codes, which were enough codes for the 26 small and 26 capital letters, plus a fairly large number of symbols and control codes. Using 1 byte (8 bits) for each character meant that the extra bit could be used for error checking using a scheme called parity. Error checking each character became more of an issue, so the American Standard Code for Information Interchange (ASCII) was extended to 8 bits, allowing the addition of 128 new codes. Characters in many of the Latin languages could be encoded using 8 bit ASCII.

12.5.4.1.2. Unicode

Computer usage has transcended national boundaries; it is now just as easy to communicate with the other side of the Earth, as to communicate within a single building. Internationalization has meant that a single server or client needs to be multilingual. Multilingual programming is now common. The ASCII code is not adequate for this task, so a multilingual code, Unicode, has superseded it. For example, the Java programming language specifies that programs be written in Unicode. Unicode uses 32 bits, and there are more than 100 000 character codes, including all common languages and many uncommon languages.

There are different ways to implement Unicode called, Unicode Transformation Format (UTF) encodings. One of the most common is a system called UTF-8, which uses a variable length coding system. Different numbers of bytes from one to four are used for different character codes. If the first byte in a code has a zero in the highest bit, then 1 byte is used. If the first byte in a code has a one in the highest bit, then subsequent bytes are used. The single byte codes are identical to the 7-bit ASCII codes. For a document that only includes the characters in the 7-bit ASCII system, encoding in ASCII and UTF-8 are identical. This results in an efficient encoding of Latin languages. In addition, any other Unicode character — Greek, Japanese or symbol — can be included occasionally in a predominantly Latin text while maintaining average coding efficiency.

12.5.4.1.3. Markup language

The difference between text editors and word processors is that the former largely edit the characters, while the latter define how the document will appear — the size of the text, the spacing, indentations, etc. For a word processor, the layout of the document as it is entered is the same as the layout when it is printed. In computer jargon, 'what you see is what you get' (WYSIWYG).

Markup has the advantage that it can be read by humans and can be edited with any text editor. Hypertext Markup Language (HTML), the language used by the World Wide Web, was originally a markup language. When HTML was first developed, text editors were used to writing it. Page layout capabilities were added, and now WYSIWYG editors are generally used. A markup language allows other types of information to be included. For example, one of the key features of HTML is that it includes hyperlinks to other HTML pages on the Internet.

12.5.4.1.4. Extensible Markup Language

Currently, a very popular and increasingly used method for encoding text information is a standard called Extensible Markup Language (XML). It provides a method of producing machine-readable textual data. For example, the Real Simple Syndication (RSS) standard is encoded with XML. A common use of the RSS standard is for web publishing. Almost all newspaper web sites use RSS to communicate with subscribers.

XML is not a markup language itself, but rather a general format for defining markup languages. It defines how the markup is written. A document is said to be 'well formed' if it follows the XML standard. If it is well formed, then the markup can be separated from the rest of the text. More information, provided by an XML schema, defines a particular markup language. The most recent version of HTML, XHTML, is fully compatible with the XML standard.

For nuclear medicine file formats, one of the key properties of XML is that it can be used to make text information readable by machine. For example, consider the following section of an XML document:

```
<patient>
    <name>
        <last>Parker</last>
        <first>Tony</fist>
    </name>
    <medical_record_number>10256892</medical_record_number>
</patient>
```

It would be straightforward for a computer to unambiguously determine the name and MRN from this document. Human readability and text editor processing of XML make it highly self-documenting. Owing to its wide acceptance in the computer world, XML is the most appropriate current format for storing non-imaging nuclear medicine information.

XML can be used to store numerical data as characters, e.g. <number>12.56</number>. The scalable vector graphics format for vector graphics is an XML language. However, XML is too inefficient for coding raster graphics image information.

12.5.4.2. Non-image information in general formats

(a) JPEG 2000: An extension of the JPEG format, JPEG 2000[1], has several very interesting capabilities. The JPEG 2000 standard is a general purpose image format; however, it was developed with medical imaging as a potential application. Since it uses wavelet-transform compression, the image quality at high compression ratios is considerably better than with DCT compression. JPEG 2000 allows multiple frames, so it could be used for cine; it has an advantage over GIF of providing good compression of information rich images such as natural scenes. However, the reason that it is included in this section is its capability to carry considerable meta-information in tight association with the image data. Although this format was defined nearly a decade ago, it has not found wide acceptance. The JPEG 2000 wavelet-transform coding is allowed in the Digital Imaging and Communications in Medicine (DICOM) standard, but mass market software, such as browsers, generally do not support this standard. It has

[1] http://www.jpeg.org/jpeg2000/

been included not as a solution that is currently practical, but rather as an example of a non-medical standard that has capabilities that could be applied to the particular requirements of nuclear medicine.

(b) Portable Document Format (PDF): PDF has not been used extensively in medicine, but it has many useful capabilities. It is intended for page layout. It can contain text, audio, images and movies. It is very widely supported, including support by all web browsers. Although uncommonly used in medicine, it has many properties that would make it an excellent format for distribution of image or cine information.

12.5.4.3. Image information

The image portion of a nuclear medicine study can be a sequence of static images, a dynamic series, a gated series, multiple dynamics series, raw data from a tomographic collection, reconstructed images, a dynamic series of tomographic collections, a dynamic series of gated tomographic collections, related tomographic datasets, curves from regions of interest, functional datasets derived from calculations on other datasets, etc. Thus, it should be clear that a rather general data format is needed. This section will present suggestions for general image format principles.

12.5.4.3.1. Types of data

The types of data in the last paragraph are not only long, but also incomplete; it is easy to think of other variations. A different data format for each type of data would add complexity without functionality. Therefore, it is logical to separate the interpretation of the data from the data. Whether the data represent a dynamic series or a set of static planar images does not have to be represented in the dataset. There is no necessity to have two dataset formats — one for a 128 × 128 image and another for a set of 128 curves with 128 points.

The interpretation of the data should be part of the non-image information, rather than part of the data. The data type should be determined by the format of the data elements. Separation of the interpretation from the data format will simplify processing and handling of the data. Adding two 128 × 128 images and adding two sets of 128 curves with 128 points is exactly the same operation; an add function should not concern itself with the interpretation of the data.

12.5.4.3.2. Data element

Several different data elements are needed. The raw data collection bins can generally be 8 or 16 bit unsigned integers, depending on whether a maximum of

255 or 65 535 counts per pixel are required. Since processed data may require a larger dynamic range, floating point or complex data may be appropriate in many circumstances. For some analysis, signed integers may be most appropriate. A region of interest can be represented as a single bit raster or with vector graphics.

The selection of a small number of data element formats would simplify the programming task. The trick is to limit the number of formats while maintaining all of the functionality. It would be logical to select some standard set of data element formats, especially a set of formats associated with a widely accepted data standard. However, it is not clear that such a standard exists. At a minimum, 16 bit integer and floating point formats are required. Probably signed and unsigned 8 and 16 bit formats should be included. It is not clear whether increasing complexity by including a vector graphics format is worth the added functionality.

12.5.4.3.3. Organization

A logical first level of organization of the image data is what will be called a ‘dataset’. A dataset is an *n* dimensional set of data in which each of the data elements is the same format, e.g. 8 bit unsigned integer, 16 bit signed integer and IEEE 32 bit floating point data. The key characteristic is that the dataset is a sequence of identical elements. The dimensions do not need to be the same; 7 sets of 100 curves with 256 points is a 7 ×100 × 256 dataset.

The dataset should be the ‘atom’ of the data; there should not be a lower level of organization. The lower levels of organization will depend on the type of data and the dimensions. For example, a 256 × 256 × 256 tomographic volume could be considered as 256 axial slices of 256 × 256, but it would be equally valid to consider that dataset as 256 coronal slices of 256 × 256. The organization of a dataset consisting of 4-D PET lines of response will depend entirely on the reconstruction algorithm. The lower levels of organization depend on the non-image information and should not be part of the image data format.

12.5.5. Common nuclear medicine data storage formats

12.5.5.1. Interfile

The Interfile format has been used predominantly in nuclear medicine. The final version of Interfile, version 3, was defined in 1992. Although Interfile has been largely replaced by DICOM, it has some interesting properties. The metadata, encoded in ASCII, are readable and editable by any text editor. The lexical structure of the metadata was well defined, so that it is computer readable.

12.5.5.2. Digital Imaging and Communications in Medicine

DICOM is the most widely accepted radiological imaging format. DICOM began as ACR-NEMA, a collaboration between the American College of Radiology and the National Electrical Manufacturers Association. Version 3 of that standard changed the name to DICOM, in part to position the standard in a more international framework. DICOM is often thought of as a file format; however, the standard covers communication more broadly. It defines a transmission protocol, a query and retrieval standard, and workflow management. Unfortunately, it is overly complex, non-self-describing and has a heavy 2-D image bias (see Section 12.6.4).

12.6. INFORMATION SYSTEM

12.6.1. Database

Databases are one of the most common applications for computers. Web sites frequently depend heavily on a database. For example, search engines provide information from a database about the web and on-line retailers make extensive use of databases both for product searches and for information about customers and orders.

12.6.1.1. Table

Almost all of the information in databases is contained in tables. A table is a simple 2-D matrix of information. The rows in a table are called records and the columns are called fields. The rows refer to individual entities. In a customer table, the rows are customers. In an order table, the rows are orders. In a patient table, the rows are patients. In a study table, the rows are the studies. The columns are the attributes of an entity. In a patient database, the columns might be first name, middle name, last name, date of birth, MRN, etc. In an administer dose table, the columns might be radioisotope, radiopharmaceutical, administered dose, time of injection, etc.

A key concept is that tables are very simple — 2-D matrixes of data. The simplicity enables reliability. Almost all of the information in the database is in a simple format that is easy to back up and easy to transfer between databases.

There is usually one field in each table that is unique for each row. That unique field can be used for lookup. As that field allows access to the record, it is called an accession number. The accession number for a patient table might be the MRN; the accession number for a study table might be the study number;

the accession number for a customer table would be the customer number. Some databases use a combination of fields as the accession number, but the basic idea is the same, there must be something unique which identifies each row.

12.6.1.2. Index

An index provides rapid access to the records (rows) of a table. The index is separate from the table and is sorted to allow easy searching, often using a tree structure. The tables themselves are usually not sorted; the records are just added one after another. The index, not the table provides organization of the records for fast access. Indexes are one of the important technologies provided by a database vendor.

Indexes can be complicated and complex. However, there is no information stored in an index. An index can be rebuilt from the tables. In fact, when transferring data to a new database, the indexes are usually rebuilt. Indexes are important for efficiency, but the only information they contain is how to efficiently access the tables.

12.6.1.3. Relation

A key element of relational databases is relations. Relations connect one table to another table. Conceptually, the relations form a very important part of a database, and provide much of the complexity. However, the relations themselves are actually very simple. A relation between the patient table and the study table might be written:

patient.MRN = study.MRN

This says that the records in the patient table are linked to the records in the study table by the MRN field.

A physician thinks of the database in terms of patients who have studies that involve radiopharmaceuticals; however, a radiopharmacist thinks of isotopes and radiopharmaceuticals. The relations facilitate these different points of view. The physician can access the database in terms of patients and the radiopharmacist can access the database in terms of radiopharmaceuticals. The same tables are used; it is just the small amount of information contained in the relations that is different.

12.6.2. Hospital information system

A hospital information system is a large distributed database. Data come from clinical laboratories, nuclear medicine and financial systems, etc.

12.6.2.1. Admission, discharge, transfer

Most hospital information systems have an admission, discharge, transfer (ADT) database. The ADT system is the master patient identification system. Other systems use the ADT system for institution-wide, reliable, coordinated identification of each patient.

12.6.2.2. Information gateway

Often, the goal in a business or a division of a larger organization is to maintain a simple business model — a small product line, etc. The simple business model is often reflected in a simple database design. Medicine is very complex and the principle of a simple, well structured design is a dream. Generally, the different departments in the hospital have incompatible databases; even within one department, there may be incompatible systems. One method of ameliorating this problem is the development of Health Level Seven (HL7), a standard message format for communicating between systems in a hospital. However, the connections between systems still differ. If there are n systems, communication between them becomes a problem that grows proportionally to n^2.

An information gateway ameliorates this problem. The only task of an information gateway is to connect systems, translating messages so that they can be understood by other systems. Each system only needs to connect to the gateway, and the gateway communicates with all of the systems in the hospital. The growth in complexity tends to increase more like n than n^2.

12.6.3. Radiology information system

The radiology information system (RIS) supports scheduling, performing, reporting procedure results and billing. When nuclear medicine is a division of radiology, the RIS usually also functions as the nuclear medicine information system. However, nuclear medicine procedures have some unique characteristics, such as studies that extend over several days, which may not be well handled by a general purpose RIS.

12.6.4. Picture archiving and communication system

The image information from the imaging equipment is usually stored in a picture archiving and communication system (PACS) that is separate from but coordinated with the radiology/nuclear medicine information system. DICOM is the predominant standard for PACSs.

12.6.4.1. Study

The top level of organization in DICOM is the study. This level of organization comes from the organization of the health care system. Health care providers request services from radiology/nuclear medicine by a request for a consultation. Imaging or another service is performed and a report, ideally including image information, is returned to the provider. Each study is linked to a single patient, but a patient can have any number of separate studies.

12.6.4.2. Sequence

Sequence is the next level of organization in DICOM. Sequence comes from a sequence of images; for a volume of image data, sequence is the top level of organization. For example, it is common to collect a sequence of axial images. A more general name for this level of organization would be dataset.

12.6.4.3. Image

Originally, DICOM used a 2-D image as the basic atom. Other data structures are composed of a number of images. A considerable amount of metadata are included with each image, defining both the structure of the image, patient information, data collection information and relation of the image to other images. Somewhat more recently, a multiframe format was defined in DICOM. One file may contain information from a volume of data, from a time series, from a gated sequence, from different photopeaks, etc. Nuclear medicine tends to use the multiframe format much more commonly than other modalities. Describing the organization of a dataset in terms of multiple images is a particularly awkward feature of DICOM.

12.6.4.4. N-dimensional data

It is unfortunate that DICOM selected the image as the basic atom of organization (see Section 12.5.4.3.3). Even before it was introduced, it was apparent that an N-dimensional data model would be more appropriate. Nuclear medicine and MRI often dealt with 1-D curves, 2-D images, 3-D volumes, 4-D gated or dynamic volumes, etc. However, radiology tended to be film based, and volume data such as CT was anisotropic, so some of the early developers had an image based orientation.

12.6.5. Scheduling

Scheduling is a much more complicated task than it may initially seem. Several appointments in different departments may need to be scheduled at the same time. Sequencing of studies may be important, e.g. thyroid imaging should not be performed in proximity to iodinated contrast usage. Some studies require a prior pregnancy test or other laboratory values. Prior data need to be made available prior to performing some studies; for example, a prior electrocardiogram should be available before a myocardial stress study.

Since scheduling needs to be coordinated between departments and make use of hospital-wide information, it is logically a function of the hospital information system. However, scheduling involves many issues that are local to the modality, for example, availability of resources such as radiopharmaceuticals, staff and equipment. Often, scheduling is local, making use of humans to provide much of the hospital-wide coordination.

The scheduling system creates a ‘worklist’, a list of studies that need to be performed. Most modern imaging equipment can use a worklist provided by the RIS. When the technologist starts an imaging study, the appropriate patient is picked from the worklist. Selecting a patient from a list results in far fewer errors than re-entering all of the demographic identifier data for each study.

12.6.6. Broker

A broker is essentially an information gateway. The term ‘information gateway’ is generally used for a hospital-wide system. The term ‘broker’ is generally used when talking about a system within a radiology department. The broker handles incompatible details related to RIS, PACS and imaging devices that are local to radiology.

12.6.7. Security

Health information is private. Although the privacy of health information is often a relatively minor concern for the general public, it is a major concern for politicians, who make rules about the security needed for medical information. The theory of secure communication over insecure channels is thought to be a solved problem. However, vigilance is necessary, because the practical implementations of the theory often have security holes that hackers can exploit. Furthermore, with aggregation of private health information, the potential extent of a security breach becomes catastrophic. Security depends not only on computer systems, but also on humans who are often the weak link.

There needs to be a balance between the damage caused due to a security breach and the expense of the security measures. Often, relatively low damage situations are addressed with overly expensive systems, especially in terms of lost productivity. The dominant effect of security should not be to prevent authorized users from accessing information. Nuclear medicine tends to be a relatively low risk environment, so in most circumstances the balance should favour productivity over security.

BIBLIOGRAPHY

HUTTON, B.F., BARNDEN, L.R., FULTON, R.R., “Nuclear medicine computers”, Nuclear Medicine in Clinical Diagnosis and Treatment (ELL, P.J., GAMBHIR, S.S., Eds), Churchill Livingston, London (2004).

LEE, K.H., Computers in Nuclear Medicine: A Practical Approach, Society of Nuclear Medicine, 2nd edn (2005).

PARKER, J.A., Image Reconstruction in Radiology, CRC Press, Boston, MA (1990).

PIANYKH, O.S., Digital Imaging and Communications in Medicine (DICOM): A Practical Introduction and Survival Guide, Springer, Berlin (2008).

SALTZER, J.H., KAASHOEK, M.F., Principles of Computer System Design: An Introduction, Morgan Kaufmann, Burlington (2009).

TODD-POKROPEK, A., CRADDUCK, T.D., DECONINCK, F., A file format for the exchange of nuclear medicine image data: a specification of Interfile version 3.3, Nucl. Med. Commun. **13** (1992) 673–699.