3

Image acquisition and representation

3.1 Sampling and quantization

Any visual scene can be represented by a continuous function (in two dimensions) of some analogue quantity. This is typically the reflectance function of the scene: the light reflected at each visible point in the scene. Such a representation is referred to as an *image* and the value at any point in the image corresponds to the intensity of the reflectance function at that point.

A continuous analogue representation cannot be conveniently interpreted by a computer and an alternative representation, the *digital image*, must be used. Digital images also represent the reflectance function of a scene but they do so in a *sampled* and *quantized* form. They are typically generated with some form of optical imaging device (e.g. a camera) which produces the analogue image (e.g. the analogue video signal discussed in the preceding chapter), and an analogue-to-digital converter: this is often referred to as a 'digitizer', a 'frame-store', or 'frame-grabber'.

The frame-grabber samples the video signal in some predetermined fashion (usually in an equally spaced square grid pattern) and quantizes the reflectance function at those points into integer values called grey-levels (see Figure 3.1). Each integer grey-level value is known as a *pixel* and is the smallest discrete accessible sub-section of a digital image. The number of grey-levels in the (equally spaced) grey-scale is called the quantization or grey-scale resolution of the system. In all cases, the grey-scale is bounded by two grey-levels, black and white, corresponding to the minimum and maximum measurable intensity respectively. Most current acquisition equipment quantizes the video signal into 256 discrete grey-levels, each of which are conveniently represented by a single byte. In certain cases, a grey-scale of -128 to $+127$ is more convenient; processed images need not necessarily represent the reflectance function and pixels may assume negative values but the grey-level can still be represented by a signed-byte integer.

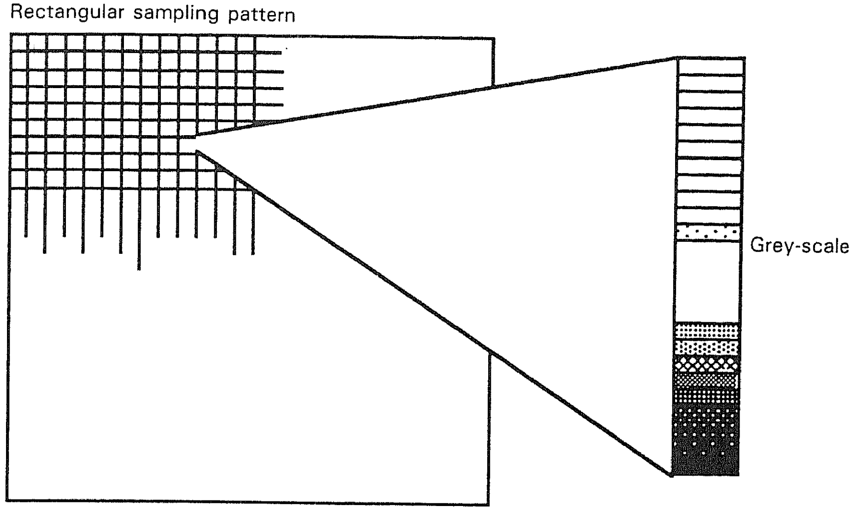


Figure 3.1 Sampling and quantization.

The sampling density, the number of sampling points per unit measure, is usually referred to as the (spatial) resolution and, since the sampling device is usually arranged as a square grid, it is measured in terms of the number of sampling elements along each orthogonal axis. This normally corresponds to the extent of the number of pixels in both the horizontal and vertical directions. Most current commercial frame-grabbers have spatial resolutions of 512 × 512 pixels. In summary, digital image acquisition equipment is essentially concerned with the generation of a two-dimensional array of integer values representing the reflectance function of the actual scene at discrete spatial intervals, and this is accomplished by the processes of sampling and quantization. Since these are fundamentally important concepts, we will look at them in a little more depth in the remainder of this section.

3.1.1 Spatial frequency and the effects of sampling

Recall that the objective of the image acquisition system (which includes the sensor sub-system) is to convert an analogue optical image to a digital image in as faithful a manner as possible. As we have seen, this is achieved by first using the sensor to sample the analogue optical image, generating an analogue video signal, and then by subsequently re-sampling the video signal and generating the digital signal. Thus, there are three factors which can limit the effective resolution and fidelity of the final digital image:

(a) the sensor sampling frequency;
(b) the bandwidth of the video signal;
(c) the sampling frequency of the analogue-to-digital converter, i.e. the frame-grabber.

We shall consider each of these in turn. First, however, we need to look a little closer at the idea of sampling an analogue image and to develop the concept of spatial frequency. We will do this intuitively at first and then we will formalize the ideas somewhat. Readers who find this area interesting should consult Duda and Hart (1973) which contains a particularly lucid account of this topic.

High-frequency signals, such as high-pitch soundwaves, periodically change their value over a very short period of *time*. Similarly, a high-frequency spatial signal, such as an image, periodically changes its value (e.g. its intensity or grey-level) over a very short *distance*, i.e. it changes abruptly from one grey-level to another. Conversely, low spatial frequencies correspond to 'slower' changes in intensity where the change occurs gradually from one position in the image to another.

To make this idea more formal, consider a spatially unbounded analogue optical image $g(x, y)$. The *Fourier transform* $G(f_x, f_y)$ of the image $g(x, y)$ is defined by:

$$\begin{aligned} G(f_x, f_y) &= \mathscr{F}(g(x, y)) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \exp[-2\pi i(f_x x + f_y y)] \, dx \, dy \qquad (3.1) \end{aligned}$$

The *inverse Fourier transform* of $G(f_x, f_y)$ is defined by:

$$\begin{aligned} g(x, y) &= \mathscr{F}^{-1}(G(f_x, f_y)) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(f_x, f_y) \exp[2\pi i(f_x x + f_y y)] \, df_x \, df_y \qquad (3.2) \end{aligned}$$

The variables f_x and f_y which identify the domain over which $G(f_x, f_y)$ is defined are known as the spatial frequencies in the x- and y-directions, respectively, and the domain is known as the *spatial frequency* domain. What do these variables and this domain represent? The integral in equation (3.2) can be viewed as a 'generalized sum' of complex exponentials, defined in terms of the spatial frequencies f_x and f_y, and in terms of the spatial coordinates x and y. Each exponential is weighted by a value given by $G(f_x, f_y)$ and, thus, equation (3.2) is an expansion (or expression) of the analogue optical function $g(x, y)$ in terms of this weighted generalized sum of exponentials. These weights are, in fact, given by equation (3.1), i.e. the Fourier transform of the image function, and will, of course, vary with image content. Since these complex exponentials can also be expressed in terms of sine functions,* a spatial frequency domain which has, for example, just two non-zero values at, say (f_{x_1}, f_{y_1}) and $(-f_{x_1}, -f_{y_1})$ corresponds to a sinusoidal variation in intensity in the spatial domain, i.e. to an image $g(x, y)$ which comprises sinusoidally undulating 'stripes' of alternating light and dark intensity. The period and orientation of these stripes depends on the exact values of f_{x_1} and

---

* $e^{i\theta} = \cos\theta + i \sin\theta$.

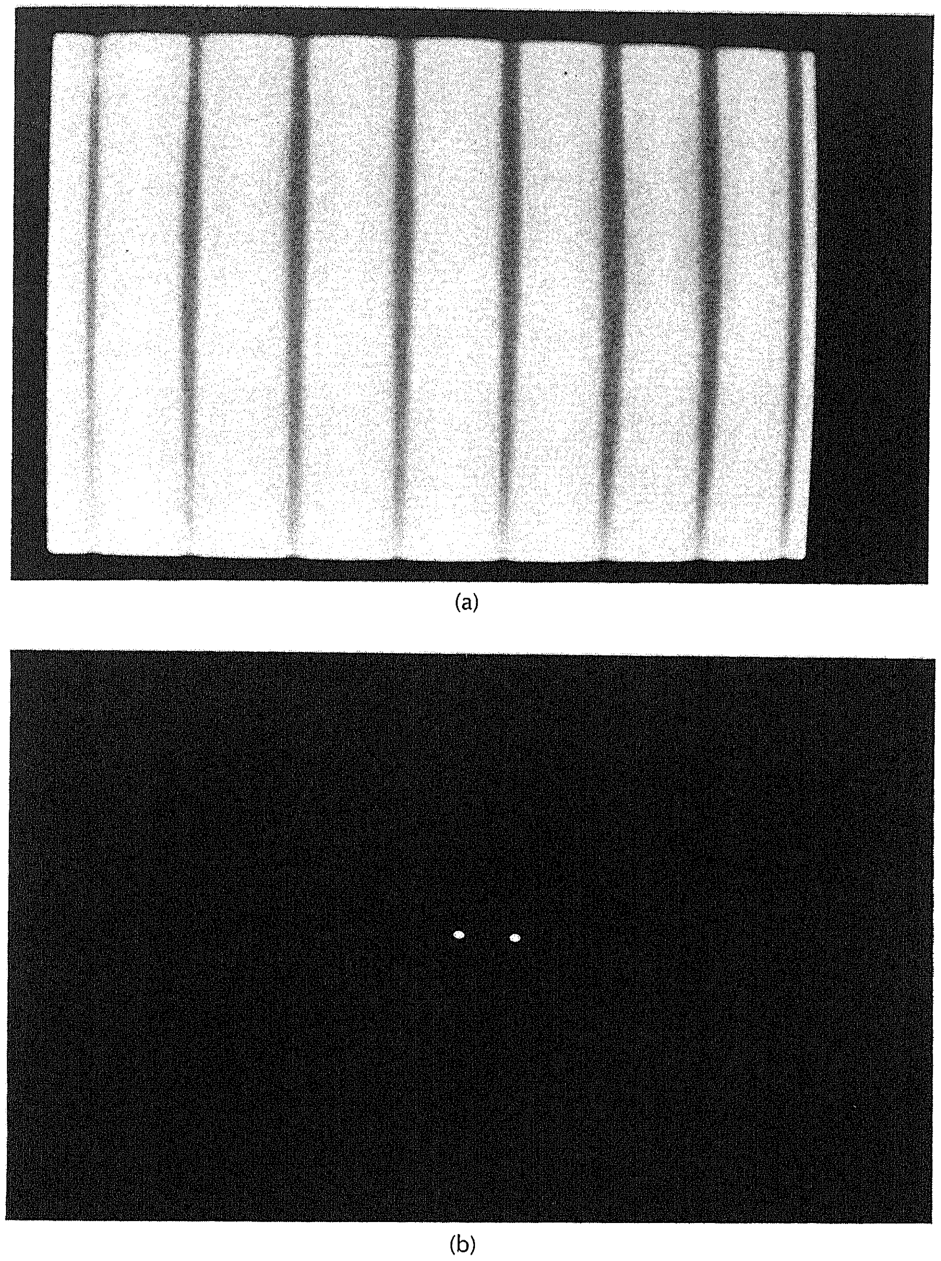


Figure 3.2 (a) An image comprising a sinusoidal variation in intensity along the x axis; and (b) its Fourier transform, comprising two spatial frequency components (f_{x_1}, f_{y_1}) and $(-f_{x_1}, -f_{y_1})$, both of which are spatial frequencies in the x direction.

f_{y_1}. Conversely, an image $g(x, y)$ which comprises a sinusoidal variation in intensity can be expressed in terms of the spatial frequencies (f_{x_1}, f_{y_1}) and $(-f_{x_1}, -f_{y_1})$; see Figure 3.2. The 'quicker' these sinusoidal variations, i.e. the greater the frequency of variation in the spatial domain, the further (f_{x_1}, f_{y_1}) and $(-f_{x_1}, -f_{y_1})$ are from the origin in the $G(f_x, f_y)$ domain. Of course, a sinusoidal variety in intensity is not a particularly common image. However, Fourier analysis tells us that more complex functions can be constructed by including more terms of varying weight in the 'generalized sum' of exponentials, i.e. *by including further spatial frequency components*. The exact weight is, again, determined by equation (3.1), i.e. the Fourier transform. An abrupt change in intensity will require the presence of a large number of terms which will correspond to high spatial frequencies, many of which are far removed from the origin of the $G(f_x, f_y)$ domain; see Figure 3.3. Thus, we now have arrived at the interpretation we required. That is: *high spatial frequencies correspond to the presence of abrupt, or sharp, changes in the intensity of the image.*

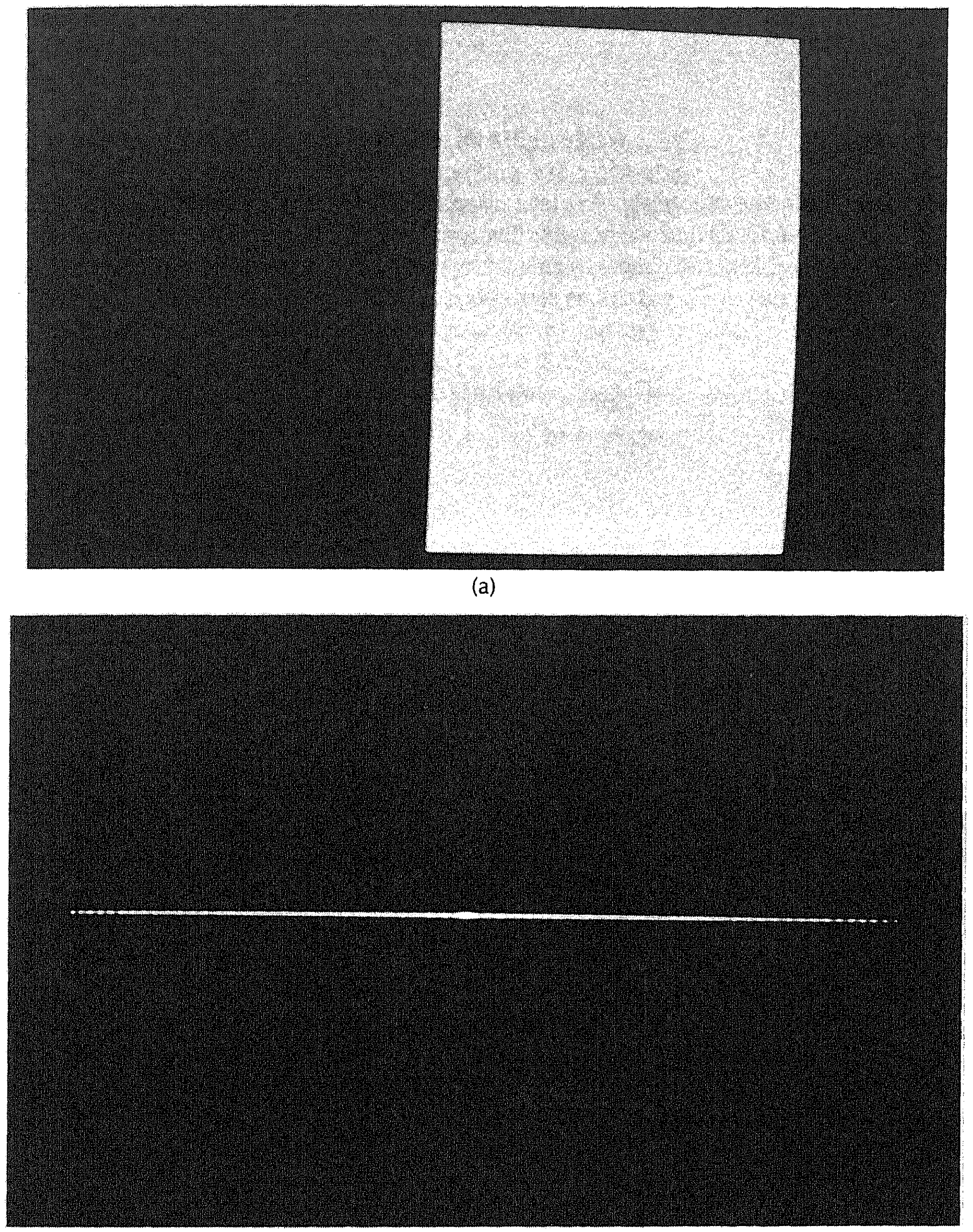
The next issue to which we turn is that of sampling. In particular, we would like to know what sampling frequency is required, i.e. how often one needs to sample in the spatial domain in order to represent an image faithfully. Shannon's sampling theorem tells us that a band-limited image (i.e. an image which does not comprise infinitely high spatial frequencies) can be faithfully represented (i.e. reconstructed) if the image is sampled at a frequency twice that of the highest spatial frequency present in the image. This sampling frequency is often referred to as the *Nyquist frequency*.

We are now in a position to return to address the three issues we raised at the beginning of this section: the effect of the sensor sampling frequency (resolution), the video signal bandwidth, and the analogue-to-digital (frame-grabber) sampling frequency on the effective resolving power of the image acquisition system.

First, we now see that a sensor which has, for example, 756 photosites in the horizontal direction, i.e. along a line, will only be capable of representing a sinusoidal variation in intensity which has a spatial frequency of 378 cycles per unit distance. A pattern of 378 alternately light and dark bars with abrupt edges, i.e. discontinuities in intensity, would obviously require a much higher sensor sampling frequency to represent the image faithfully.

Second, the resolution of a television picture is also limited by the number of lines and the frequency bandwidth of the video signal. We saw in Chapter 2 that the line frequency for the CCIR standard video signal is 15 625 Hz. In addition, the nominal bandwidth of the CCIR standard is 5.0 MHz, meaning that a signal can transmit a video image with five million periodic variations in the signal (brightness) levels. This results in an absolute maximum of $5 \times 10^6 \div 15\,625 = 320$ periodic (or sinusoidal) variations per line, that is, the maximum spatial frequency which can be faithfully represented by a video signal is 320 cycles per line.

Third, a frame-grabber which has, for example, a sampling frequency of 512 pixels in the horizontal direction, will only be capable of faithfully representing a sinusoidal variation in intensity which has a spatial frequency of 256 cycles per unit



(a)

(b)

Figure 3.3 (a) An image comprising a step discontinuity in intensity along the x axis; and (b) its Fourier transform, exclusively comprising spatial frequency components f_x, i.e. spatial frequencies in the x direction.

distance. Again, a pattern of 256 alternately light and dark bars with abrupt edges would require a much higher sampling frequency to represent the image faithfully.

3.2 *Inter-pixel distances*

As mentioned in the preceding chapter, video signals assume an aspect ratio of 4:3 (horizontal-to-vertical) and we noted in Section 3.1 that, although framestores with 4:3 aspect ratios are becoming available, they normally use a square aspect ratio of 1:1. The aspect ratio mis-match has serious repercussions for gauging or measuring functions of an inspection system: the rectangular picture is being squeezed into a square image and, hence, the effective distance between horizontal pixels is $\frac{4}{3}$ times greater than that between vertical neighbours (see Figure 3.4).

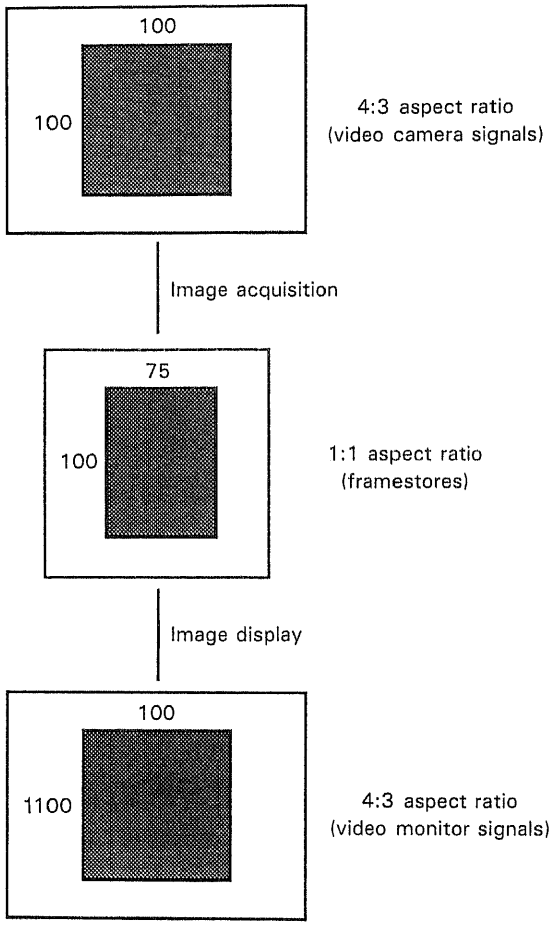


Figure 3.4 Distortion arising when acquiring a video signal with a conventional (square) framestore.

The situation becomes even more unsatisfactory when one considers the distance between a pixel and its diagonal neighbour. While most discussions of inter-pixel distances usually assume the diagonal inter-pixel interval to be $\sqrt{2}$ (i.e. the length of the hypotenuse completing the right-angle triangle formed by the horizontal inter-pixel interval of length 1 and the vertical inter-pixel interval of length 1). However, if we are working with a framestore which has a square aspect ratio and has a video signal which as a 4:3 aspect ratio, then the diagonal inter-pixel interval is, in fact, $\sqrt{\frac{5}{3}}$, i.e. the length of the hypotenuse completing the right-angle triangle formed by the horizontal inter-pixel interval of length $\frac{4}{3}$ and the vertical inter-pixel interval of length 1 (see Figure 3.5).

Unfortunately, this is not the complete story. The CCIR standard stipulates that a picture comprises 625 lines. However, only 576 of these carry visual information while the remainder are used for other purposes. Framestores with a vertical resolution of 512 pixels (i.e. 512 lines) *do not* capture all 576 of these lines of video; they only capture the first 512 of them. This introduces a further distortion, resulting in an effective CCIR aspect ratio of $4{:}(3 \times \frac{512}{576}) = 4{:}2.66$ or 3:2. The effective vertical, horizontal and diagonal inter-pixel distances are thus 1, $\frac{3}{2}$, $\sqrt{\frac{13}{4}}$.

3.3 *Adjacency conventions*

There is yet another problem which is associated with the representation of digital images: the exact spatial relationship of one pixel with its neighbours. In effect, this problem is one of defining exactly which are the neighbours of a given pixel. Consider the 3×3 neighbourhood in an image (shown in Figure 3.6) where the

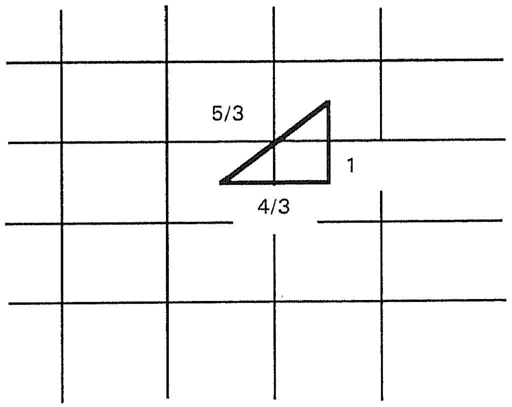
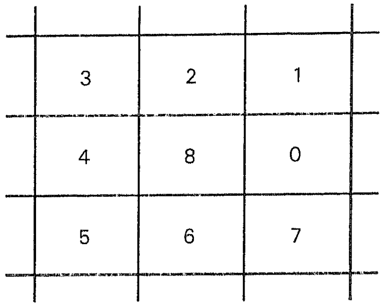


Figure 3.5 Inter-pixel distances.



3	2	1
4	8	0
5	6	7

Figure 3.6 A 3 × 3 pixel neighbourhood.

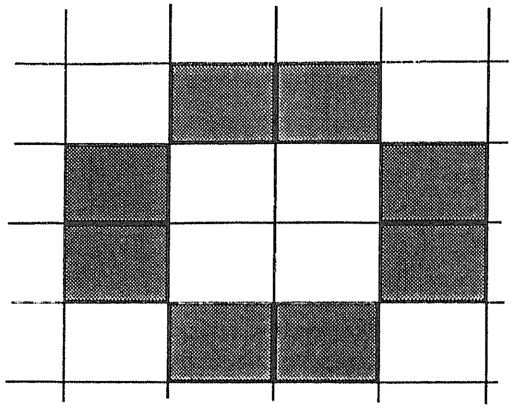


Figure 3.7 Adjacency conventions: a dark doughnut on a white table.

pixels are labelled 0 through 8; which pixels does pixel 8 touch? One convention, called 4-adjacency, stipulates that pixel 8 touches (i.e. is connected to) pixels 0, 2, 4, and 6, but does not touch (i.e. is not connected to pixels 1, 3, 5, and 7). An alternative convention, called 8-adjacency, stipulates that pixel 8 is connected to all eight neighbours.

Adopting either convention universally can, however, lead to difficulties. Figure 3.7 shows parts of a digital image depicting, in extremely simplified form, a dark doughnut on a white table. If we apply the 4-adjacency convention, then we have an obvious problem: there are four 'doughnut segments' (two vertical and two horizontal) but none of the segments is touching: the segments are not connected. Applying the 8-adjacency convention, the segments are now connected in a ring (as we would expect) but now so too is the inside of the doughnut connected to the outside: a topological anomaly.

In themselves, neither convention is sufficient and it is normal practice to use both conventions: one for an object and one for the background on which it rests. In fact, this can be extended quite generally so that the adjacency conventions are applied alternately to image regions which are recursively nested (or embedded) within other regions as one goes from level to level in the nesting. Thus, one would apply 4-adjacency to the background region, 8-adjacency to the objects resting on the background, 4-adjacency to holes or regions contained within the objects, 8-adjacency to regions contained within these, and so on. This convention means that one never encounters topological anomalies such as the one described in the above example. Because the 8-adjacency convention allows diagonally connected neighbours, measurements of object features (e.g. its perimeter) will be more faithful with 8-adjacency. Consequently, if it is possible to stipulate which convention is to be applied to a particular object of interest, one should opt for the 8-adjacency convention.

The preceding discussion has been developed in the context of the acquisition of digital images from area cameras. As we have seen, however, line-scan cameras are extremely important imaging devices, particularly for high-resolution gauging applications. The same sampling and quantization processes must also be used with these linear array cameras, the only difference in this case is that only one line of video information needs to be digitized. Similarly, the discussion of adjacency conventions and inter-pixel distances applies equally. Image acquisition equipment for line-scan sensors will often be custom-designed for a particular application and configuration, although it is worth reiterating the point made earlier that matched line-scan cameras and digitizers and, indeed, line-scan digitizers are now appearing on the market. These line-scan digitizers are, in fact, general purpose devices which can deal with many different scan rates. They are often referred to as slow-scan digitizers. This is an unfortunate misnomer: they are really variable-scan digitizers and can deal with extremely high data rates.

3.4 Image acquisition hardware

A typical machine vision system can be configured in two ways: by building it yourself using off-the-shelf equipment or by buying a complete turnkey system. The former alternative is very cost-effective but a significant amount of work needs to be done to integrate the system, ensuring that the image acquisition and processing devices are correctly matched with the host computer and can be controlled by it. The latter alternative is more expensive but, depending on the application, turnaround time on development should be much reduced.

There are two mutually dependent components to be chosen when configuring a machine vision system with off-the-shelf equipment: the CPU and the framestore. Most of the commercial framestores are dedicated to just a few of the more popular computer buses, in particular, the VME bus, the Multibus, the Q-Bus, the IBM PC bus, the Nu-bus (Apple Macintosh II), and to some extent the IBM MicroChannel Architecture used in the newer PS-2 models. Each bus enforces certain restrictions on the way the framestore is used and most of the functional differences are related to the bus structure and the available support equipment (i.e. the framestore sister boards).

Table 3.1 Overview of commercial frame-grabber hardware

Feature	*ITI*[a] *PCVision Plus*	*ITI Series 100*	*ITI 150/151 Family*	*ITI 200 Family*	*Matrox MVP-AT MVP-NB*	*Datacube Max Video Family*	*DT*[b] *2851/ 2858*	*DT 2603*	*DT 2861*	*DT 225*
Frame buffers	2	up to 4	2 per board[c]	up to 4 per board[c]	4	Many[c]	2	1	16	1
Spatial resolution	512 × 512	512 × 512	512 × 512	512 × 512	512 × 512	512 × 512	512 × 512	256 × 256	512 × 512	768 × 512
Bus compatibility	PC-AT	PC-AT VME Multibus Q-bus	VME 151: VME-AT adaptor	Q-bus	PC-AT	VME	PC-AT Q-bus VME	PC-AT	PC-AT	Nu-bus
RS-170	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CCIR	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Grey levels	256	256	256	256	256	256	256	64	256	256
Dedicated Graphics plane	0	4	—	—	4	Several	0	2	0	0
I/O mapped	No	Yes	—	Yes	—	—	Yes	—	—	—
Memory mapped	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Input LUTS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo-colour	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Pan/scroll/zoom	Yes	Yes	Yes[c]	Yes[c]	Yes	Yes[c]	—	No	—	No
Dedicated video bus	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No
Video inputs	2	3	4	4	1	8	1	1	1	4
4:3 aspect ratio compensation	Yes	Yes	No	No	No	No	No	No	No	Yes

Feature	*ITI*[d] *PCVision Plus*	*ITI Series 100*	*ITI 150/151 Family*	*ITI 200 Family*	*Matrox MVP-AT MVP-NB*	*Datacube Max Video Family*	*DT*[e] *2851/ 2858*	*DT 2603*	*DT 2861*	*DT 225*
Convolution	No	Binary images	Yes[c]	Yes[c]	Yes 3 × 3 400ms	Yes[c] up to 8 × 8 in realtime	Yes not in realtime	No	Yes not in realtime	No
Erosion	No	Binary images	Yes[c]	Yes[c]	Yes	Yes[c]	Yes	No	Yes	No
Dilation	No	Binary images	Yes	Yes[c]	Yes	Yes[c]	Yes	No	Yes	No
Sister boards	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	No
Real-time ALU operations	No	No	Yes[c]	Yes	Yes	Yes[c]	Yes	No	Yes	No
Histogram	No	No	Yes[c]	No	Yes	Yes[c]	Yes	No	—	No
Feature[f] extraction	No	No	Yes[c]	No	No	Yes[c]	No	No	No	No
Area of interest	No	No	Yes[c]	Yes[c]	Yes	Yes[c]	No	No	No	No
Prototyping board	No	No	Yes	No	No	Yes[c]	No	No	No	No
Variable Scan interface	No	No	Yes	No	No	Yes[c]	No	No	Yes	No

---

[a] Imaging Technology Corporation.
[b] Data Translation Corporation.
[c] Functionality provided by sister board.
[d] Imaging Technology Corporation.
[e] Data Translation Corporation.
[f] A feature in this context has an extremely restricted meaning: it is a region having a specified grey-level.

It would not be appropriate to enter into a discussion of different CPUs here; suffice it to say that most microprocessors are supported by VME bus and Multibus boards (e.g. 68030, 80386, 32032), while Q-bus computers are generally MicroVax based. The IBM PC AT bus is supported by IBM's PC-AT and the multitude of PC clones (most of which are now significantly faster than the AT).

Once an image has been digitized, it is stored on the frame-grabber in memory as a square two-dimensional array. It may now be accessed, processed, and analysed by the host computer. However, it is frequently desirable to process the digital image somewhat before the computer ever uses the information, mainly because the resultant image may be more amenable to subsequent analysis and it may take the computer an inordinately long time to perform such processing. This pre-processing can frequently be accomplished in one frame-time (40 ms) by hardware resident on the frame-grabber or by sister boards for the frame-grabber. This rate of processing is often called real-time image processing, as the image data can be processed at the same rate at which the camera system generates it. Recall, however, the comment in Chapter 1 that a more pragmatic definition of real-time in machine vision is simply that the vision process should keep up with production rates, i.e. if the vision system produces the information as quickly as it is required, it runs in real-time.

To provide the reader with some idea of the capabilities that are provided by some commercial frame-grabbers and sister boards, a brief summary of the main characteristics is given in Table 3.1. Unfortunately, many of the capabilities that are cited refer specifically to image processing techniques which are not covered until the next chapter. To describe them here, however briefly, would be pre-emptive and you should refer again to this section after having read Chapter 4.

Custom commercial systems, which can be bought complete with processing and analysis software, and subsequently programmed to accomplish a required function, tend to be significantly more expensive than a system configured from off-the-shelf equipment. Since such systems are by their very nature unique, a direct comparison is very difficult. However, a brief synopsis of some systems that are available is included so that the reader can form some idea of their capabilities.

Universal Instruments Corp. market several printed circuit board (PCB) inspection systems. Model 5511 exploits four high-resolution CCD cameras, each of which views a small portion of the board from a different angle in order to facilitate thorough inspection. The complete board is inspected by moving the PCB on an X–Y table. It uses high-speed strobed light emitting diode (LED) illumination to allow it to take images while the PCB is being moved on the X–Y table. This increases the overall speed of the system quite significantly since it removes the acceleration and deceleration time between snapshots. The quoted inspection performance is an impressive twelve square inches per second. The 5515 is based on the Motorola 68020 (nominally running at 20 MHz) and can use CAD data to drive the inspection strategy.

IRI (International Robomation Intelligence) Ltd offer a wide variety of systems, from development vision systems with imaging resolutions of 256 × 256 and 512 × 512, to fully fledged turnkey PCB inspection systems. The development systems are typically based on Motorola 680XX microprocessors and incorporate a real-time operating system and an extensive library of image processing functions. It should be emphasized that this library of functions is a collection of *image processing* routines, rather than *image analysis* routines. The distinction will be discussed in detail in the next chapter but suffice it for the present to note that a significant amount of software development will be required in most cases before a fully operational target system can be configured and incorporated in a manufacturing environment. A typical IRI development system, the SD512 Vision System, features 512 × 512 × 8 bit resolution, and can store eight images simultaneously. It can interface to normal video cameras, to line-scan cameras, and to special-purpose video cameras such as the Videk MEGAPLUS 1320 × 1035 sensor. Optional array processors are available to implement convolution and algebraic operations in near real time (250 million operations per second). A development system such as this would normally incorporate a MC68020 25 MHz host processor, a MC68851 memory management unit, 8 Mbytes of memory, a convolution processor, a correlation processor, CCD camera, colour monitor, terminal, 56 Mbyte hard disk, eight camera ports, and 2 Mbytes of image memory. Complete turnkey systems, e.g. the IRI PCB INSPECTOR would typically cost between two and three times the cost of a development system.

In a similar vein to IRI Ltd, Computer Recognition Systems Ltd, in the United Kingdom, offer a mix of configured development systems and application engineering expertise and have developed machine vision systems for several industries. A typical CRS development system comprises at least a VMEbus workstation (MC68020, 1 Mbyte program RAM, 40 Mbyte hard disk, 5.25″ floppy disk drive, Idris operating system, Pascal, and C), two 512 × 512 × 8 bit image memory, 8 bit resolution frame-grabber, extensive image processing algorithm library and development software, edge detector (simple 3 × 3 masks), and four video inputs.

Note that these summaries, and indeed Table 3.1, are provided for illustration only; while every effort has been made to ensure that this information is correct, manufacturers are continually upgrading their products and specific models often re-appear with much enhanced capabilities. As such, no responsibility can be accepted for errors in the functional specifications or prices. You should contact the vendor to ensure that the information is accurate and up-to-date.

3.5 Speed considerations

There are several issues which must be addressed when evaluating vision systems and their potential processing speed. The first of these is obviously the processing power of the host computer. An Intel 80286-based PC-AT will operate about three times faster than an 8086-based PC, while an 80386-based system can deliver nearly ten times the power of a PC. A DEC MicroVAX will outperform an LSI-11 or

PDP-11. Many of the newer, more powerful VME-bus systems are based on Motorola's 68030; but you should be sure to compare the processor clock frequencies and memory access times of otherwise similar CPU boards.

No matter how fast the CPU is, if it can't communicate effectively with the framestore then the effective image acquisition time will become significant, i.e. the image may be grabbed in $\frac{1}{25}$th second but it may take nearly one second to transfer it to the host. In general, memory-mapped framestores are the most efficient since transfer may not be necessary; this is the distinct advantage of VME systems, such as some of those marketed by Datacube and Imaging Technology. Many boards are input/output (I/O) mapped, however, and image transfer must take place pixel by pixel; some boards attempt to alleviate this bottleneck by offering DMA (direct memory access) transfer capabilities.

Most microprocessor systems do not support floating point arithmetic directly (the Inmos Transputer is one exception) and, if your application requires a significant number of floating point operations, this may be another computational bottleneck. Even if you can get a floating point co-processor, it is best to adhere to integer arithmetic wherever possible when writing image analysis software.

As is evident from the survey of imaging systems in the preceding section, most framestores facilitate simple image manipulation (such as contrast stretching, thresholding, trivial segmentation: see next chapter for detailed explanation) through the use of look-up tables (LUT). Each LUT will have an entry for every grey-level that the system can deal with, typically 256 of them. This entry corresponds to the value to which this grey-level is to be changed. When digitizing the video signal each incoming pixel is checked against the LUT, and the table value, not the digitized value, is stored in the frame buffer.

Most vendors now offer ALU (arithmetic logic unit) boards which contain high-speed pipe-line processors. Basic operation includes 8 bit 2's complement, logical AND, OR, XOR, multiplication, and 16 bit addition. A full frame can usually be processed in one frametime (40 ms). These boards are intended to provide real-time frame summation, thresholding, contrast enhancement, and relatively fast edge enhancement and filtering. If real-time filtering, such as edge detection, is required then sister boards can often be deployed. It is important to note that these systems generally utilize their own high-speed video bus to effect communication between the frame-grabber and the sister processing boards.

As a concluding note, remember that the application requirements will dictate the effective pixel throughout the inspection or robot vision system, which, in turn, will dictate the required architecture and whether or not the application is feasible. If real-time response is required, as it normally will be, this does not necessarily mean that it must process the part in, say, 10 ms; it means that the machine vision system must not delay the overall production system. If extremely fast on-the-fly processing is required then one may have to restrict the functionality of the system somewhat to ensure a feasible solution or, alternatively, one may have to deploy dedicated hardware to implement either the initial processing or subsequent analysis, or both.

Exercises

1. What compatibility problems would you encounter in measuring the perimeter of objects when the alternate 8-adjacency/4-adjacency convention is used?
2. What do you understand by the terms 'quantization resolution' and 'sampling density'? Identify two adjacency conventions and discuss the merits of each.
3. What adjacency convention would you choose in an application to measure visually the perimeter of coastlines in ordnance survey maps using a conventional 4:3 aspect ratio CCIR camera and a 1:1 aspect ratio framestore? How would you ensure that your measurement is as accurate as possible, given a fixed field of view?
4. What problems do you encounter when attempting to use line-scan cameras?
5. Dedicated video buses are used to alleviate the communication bottleneck between frame-grabbers and their sister boards. Why?

References and further reading

Agin, G. 1975 *An Experimental Vision System for Industrial Application*, SRI International, Technical Report No. 103.

Agin, G. 1977 *Servoing with Visual Feedback*, SRI International, Technical Report No. 149.

Agin, G. 1980 'Computer vision systems for industrial inspection and assembly', *Computer*, Vol. 13, No. 5.

Beedie, M. 1986 'Image IC detects edges in real time', *Electronic Design*, May, pp. 50–8.

Duda, R.O. and Hart, P.E. 1973 *Pattern Classification and Scene Analysis*, Wiley, New York.

Dunbar, P. 1986 'Machine vision', *Byte*, Vol. 11, No. 1, pp. 161–73.

Giordano, A., Maresca, M., Sandini, G., Vernazza, T. and Ferrari, D. 1985 *A Systolic Convolver for Parallel Multiresolution Edge Detection*, Internal Report, DIST, University of Genoa.

Giordano, A., Maresca, M., Sandini, G., Vernazza, T. and Ferrari, D. 1987 'VLSI-based systolic architecture for fast Gaussian convolution', *Optical Engineering*, Vol. 26, No. 1, pp. 63–8.

Healy, P. and Vernon D. 1988 *Very Coarse Granularity Parallelism: Implementing 3-D Vision with Transputers*, Proc. Image Processing '88, Blenheim Online Ltd., London, pp. 229–45.

Li, H.F., Tsang, C.M. and Cheung, Y.S. 1983 'A low-cost real-time imaging and processing system, *Software and Microsystems*, Vol. 2, No. 5, pp. 121–9.

Plessey Semiconductors Ltd 1986 *PDSP16401 2-Dimensional Edge Detector*, Preliminary product information.

Selfridge, P.G. and Mahakian, S. 1985 'Distributed computing for vision: architecture and a benchmark test, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI/7, No. 5, pp. 623–6.