CHAPTER 16

Video transmission and storage

16.1 Introduction

This chapter considers the means of encoding, transmitting, storing and displaying a video picture by both traditional (analogue) and modern (digital) techniques. Here we aim to give only an introduction to the subject; for more advanced and detailed information, the reader is referred to one of the many specialist textbooks, e.g. [Grob, Lenk].

The fundamental requirement of any video display system is the ability to convey, in a usable form, a stream of information relating to the different instances of a picture. This must contain two basic elements, namely some description of the section of the picture being represented, e.g. brightness, and an indication of the *location* (in space and time) of that section. This implies that some *encoding* of the picture is required.

There are many different approaches to the encoding problem. We shall look at some of the common solutions which have been adopted, although other solutions also exist. Initially, we consider analogue based solutions, which transfer and display video information in real time with no direct mechanism for short term storage. Later digitally based solutions, which allow for storage of the video image sequence, are considered. To date, however, most digital video sequences are currently displayed using analogue techniques similar to those employed in current domestic television receivers.

The problem of representing a small section of an image (often called a picture element or pixel) is solved by different encoding mechanisms in different countries. All essentially separate the information contained in a pixel into black and white (intensity) and colour components. The pixel, although normally associated with digital images, is relevant to analogue images as it represents the smallest independent section of an image. The size of the pixel limits *resolution* and, therefore, the quality of the image. It also affects the amount of information needed to represent the image. For real time transmission this in turn determines the bandwidth of the signal conveying the video stream (see Table 1.2 and Chapter 9 on information theory).

The problem of identifying the location of a pixel is addressed by allowing the image to be represented by a series of lines one pixel wide, scanned in a predetermined manner across the image, Figure 16.1. The pixels are then transmitted serially, special pulse sequences being used to indicate the start of both a new line and a new picture or image frame according to the encoding scheme used (see later Figure 16.4). Note that these systems rely upon the transmitter and receiver remaining synchronised. In the UK, the encoding mechanism used for TV broadcasting is known as PAL (phase alternate line), in the USA NTSC (National Television Standards Committee) is used, while in France SECAM (système en couleurs à mémoire) has been adopted.

16.2 Colour representation

A pixel from a colour image may be represented in a number of ways. The usual representations are:

1. An independent intensity (or luminance) signal, and two colour (or chrominance) signals normally known as hue and saturation.
2. Three colour signals, typically the intensity values of red, green and blue, each of which contains part of the luminance information.

In the second technique a white pixel is obtained by mixing the three primary colours in appropriate proportions. The colour triangle of Figure 16.2 shows how the various colours are obtained by mixing. Figure 16.2 also interprets, geometrically, the hue and saturation chrominance information. The hue is an angular measure on the colour triangle whilst the proportion of saturated (pure) colour to white represents radial distance. In practice we also need to take account of the response of the human eye which varies with colour or wavelength as shown in Figure 16.3. Thus for light to be interpreted as white we actually need to add 59% of green light with 30% of red and 11%

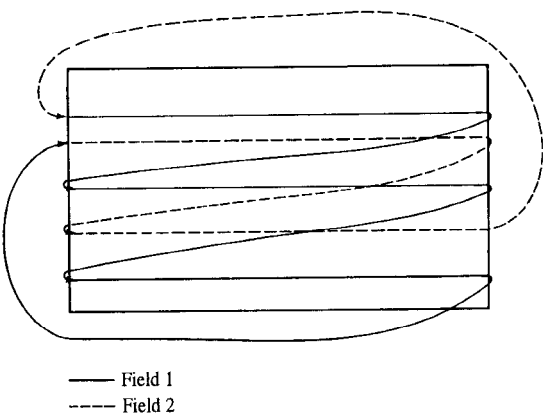

Figure 16.1 *Line scanning TV format with odd and even fields.*

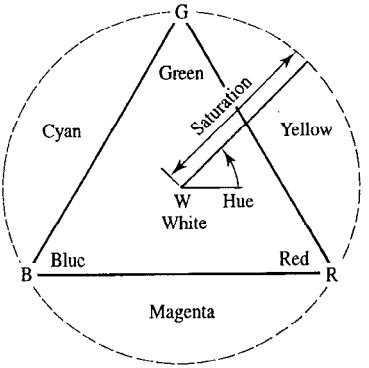

Figure 16.2 *Colour triangle showing hue and saturation.*

of blue light. The luminance component Y is thus related to the intensity values of the red (R), green (G), and blue (B) contributions by the following approximate formula:

$$Y = 0.3R + 0.59G + 0.11B \tag{16.1}$$

In practice, luminance and colour information are mathematically linked by empirical relationships. The principal benefit of separating the luminance and chrominance signals is that the luminance component only may then be used to reproduce a monochrome version of the image. This approach is adopted in colour television transmission for compatibility with the earlier black and white TV transmission system.

The theory governing the production of a range of colours from a combination of three primary colours is known as additive mixing. (This should not be confused with subtractive mixing as used in colour photography.) It is possible for full colour information to be retrieved if the luminance and two colours are transmitted. In practice, colour/luminance difference signals (e.g. R – Y) are transmitted, and these are modified to fit within certain amplitude constraints. The colour difference signals, or colour separated video components, U and V, are:

$$U = 0.88(R - Y) \tag{16.2}$$

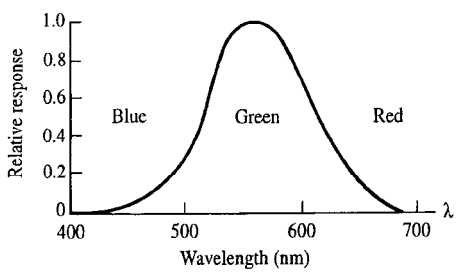

Figure 16.3 *Response of the human eye to colour.*

and

$$V = 0.49(B - Y) \tag{16.3}$$

Most video cameras and cathode-ray tube (CRT) displays produce an image in the so-called RGB format, i.e. using three signals of mixed intensity and colour information. This is transformed into YUV format before transmission and reformed into an RGB representation for colour display.

16.3 Conventional TV transmission systems

16.3.1 PAL encoding

The PAL encoding mechanism, as employed in the UK, provides for both colour information transmission and regeneration of a scanned image after display synchronisation. The latter is achieved by transmitting the image as a sequence of lines (625 in total) which are displayed to form a sequence of frames (25 per second), Figure 16.1. The system is basically analogue in operation, the start of line and start of frame being represented by pulses with predefined amplitudes and durations.

Line scanning to form the image is illustrated in Figure 16.1. In PAL systems, a complete image, or frame, is composed of two fields (even and odd) each of which scans through the whole image area, but includes only alternate lines. The even field contains the even lines of the frame and the odd field contains the odd lines. The field rate is 50 fields/s with 312½ lines per field and the frame rate is 25 frames/s.

The line structure is shown in Figure 16.4. Figure 16.4(a) shows the interval between fields and Figure 16.4(b) shows details of the signal for one scan line. This consists of a line synchronisation pulse preceded by a short duration (1.5 μs) period known as the front porch and followed by a period known as the back porch which contains a 'colour burst' used for chrominance synchronisation, as described shortly. The total duration of this section of the signal (12 μs) corresponds to a period of non-display at the receiver and is known as 'line blanking'. A period of active video information then follows where the signal amplitude is proportional to luminous intensity across the display and represents, in sequence, all the pixel intensities in one scan line. Each displayed line at the receiver is therefore of 52 μs duration and is repeated with a period of 64 μs to form the field display.

625 lines are transmitted over the two fields of one frame. However, only 575 of these contain active video. The non-active lines contain field synchronisation pulses, and also data for teletext type services. The video information carried by a PAL encoded signal is contained within a number of frequency bands, although to save bandwidth, the chrominance information is inserted into a portion of the high frequency section of the luminance signal, Figure 16.4(c). This band is chosen to avoid harmonics of the line scan frequency which contain most of the luminance energy. Fortunately the colour resolution of the human eye is less than its resolution for black and white images and hence we can conveniently transmit chrominance information as a reduced bandwidth signal within the

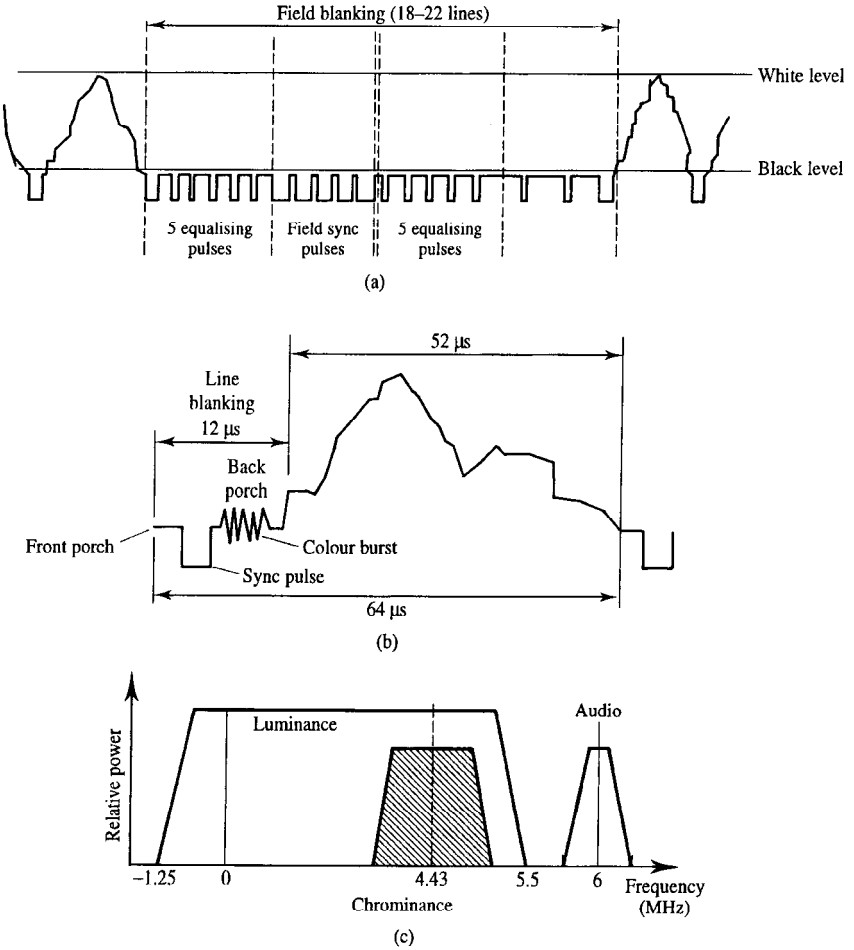

Figure 16.4 *TV waveform details: (a) field blanking information at the end of a frame; (b) detail for one line of video signal; (c) spectrum of the video signal.*

luminance spectrum.

The chrominance information is carried as a quadrature amplitude modulated signal, using two 4.43 MHz (suppressed) carriers separated by a 90° phase shift, to carry the colour difference signals of equations (16.2) and (16.3). This is similar to QAM modulation described in Chapter 11, but with analogue information signals. If the carrier frequency is f_c (referred to as the colour sub-carrier frequency) then the resulting colour signal S_c is:

$$S_c = \mathrm{U}\cos(2\pi f_c t) + \mathrm{V}\sin(2\pi f_c t) \tag{16.4}$$

The resulting chrominance signal bandwidth is 2 MHz. This is sometimes called the YIQ

signal. Figure 16.4 also shows an additional audio signal whose sub-carrier frequency is 6 MHz.

In order to demodulate the QAM chrominance component, a phase locked local oscillator running at the colour sub-carrier frequency f_c is required at the receiver. This is achieved by synchronising the receiver's oscillator to the transmitted colour sub-carrier, using the received 'colour burst' signal, Figure 16.4(b).

Since the chrominance component is phase sensitive, it will be seriously degraded by any relative phase distortion within its frequency band due, for example, to multipath propagation of the transmitted RF signal. This could lead to serious colour errors but the effect is reduced by reversing the phase of one of the colour difference components on alternate line scans. Thus the R – Y channel is reversed in polarity on alternating line transmissions to alleviate the effects of differential phase in the transmission medium.

The PAL signal spectrum shown in Figure 16.4(c) is the baseband TV signal and if RF television transmission is required, this must be modulated on to a suitable carrier and amplified. Typical UHF carrier frequencies extend to hundreds of MHz with power levels up to hundreds of kW (even MW in some cases). Terrestial TV channels 21 to 34 lie between 471.25 and 581.25 MHz and channels 39 to 68 fall between 615.25 and 853.25 MHz, Table 1.4. Satellite TV occupies a band at 11 GHz with 16 MHz interchannel spacings.

16.3.2 PAL television receiver

Figure 16.5 shows a simplified block diagram of the main functional elements of a colour television receiver. The RF signal from the antenna or other source is selected and

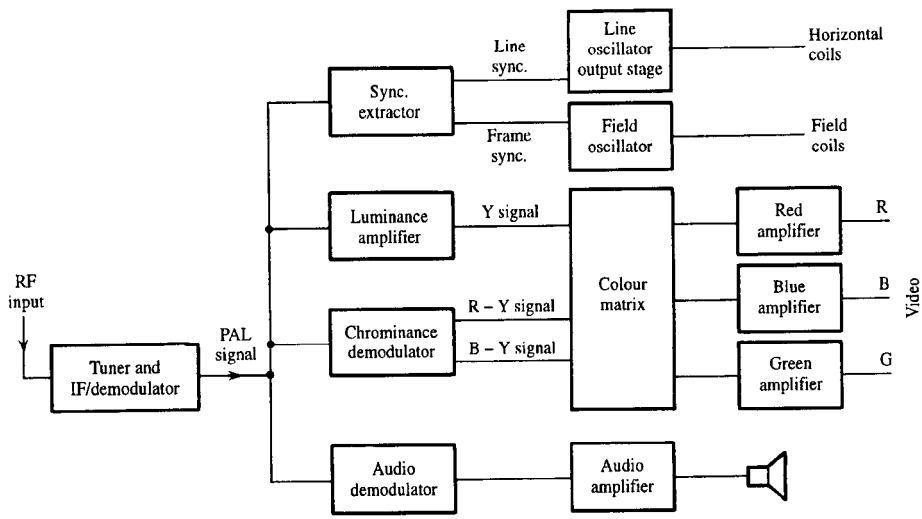

Figure 16.5 *Simplified block diagram of colour TV receiver.*

amplified by the tuner and IF (intermediate frequency) stages and finally demodulated to form a baseband PAL signal. Four information signals are then extracted: the audio, luminance, chrominance, and synchronisation signals.

The chrominance signal is demodulated by mixing with locally generated in phase and quadrature versions of the colour sub-carrier. The resulting colour difference and luminance signals are then summed in appropriate proportions and the R, G and B output signals amplified to sufficient levels to drive the CRT video display.

The x and y deflections of the electron beams are produced by magnetic fields provided by coils situated on the outside of the CRT. These coils are driven from ramp generators, synchronised to the incoming video lines such that the beam is deflected horizontally across the display face of the tube from left to right during the active line period, and vertically from top to bottom during the active field period.

16.3.3 Other encoding schemes

The NTSC system has many fundamental similarities to PAL, but uses different transmission rates. For example, the frame rate is 30 Hz, and the number of lines per frame is 525. PAL is, in fact, an enhancement of the basic principles of the NTSC system. The main difference between the two is in the phase distortion correction provided by PAL. This is not present in the NTSC system, and hence NTSC can be subject to serious colour errors under conditions of poor reception. Any attempted phase correction is controlled entirely by the receiver, and most receivers allow the viewer to adjust the phase via a manual control. The overall spectral width of NTSC is somewhat smaller than PAL.

The SECAM system uses a similar frame and line rate to PAL and, like PAL, was developed in an attempt to reduce the phase distortion sensitivity of NTSC. Essentially, it transmits only one of the two chrominance components per line, switching to the second chrominance component for the next line.

Transmission of digitally encoded video images is currently being developed and implemented. Some of the basic digital techniques are discussed in later sections.

16.4 High definition TV

16.4.1 What is HDTV?

High definition television (HDTV) first came to public attention in 1981, when NHK, the Japanese broadcasting authority, first demonstrated it in the United States. HDTV [Prentiss] is defined by the ITU-R study group as:

> 'A system designed to allow viewing at about three times the picture height, such that the system is virtually, or nearly, transparent to the quality or portrayal that would have been perceived in the original scene ... by a discerning viewer with normal visual acuity.'

HDTV proposals are for a screen which is wider than the conventional TV image by about 33%. It is generally agreed that the HDTV aspect ratio will be 16:9, as opposed to the 4:3 ratio of conventional TV systems. This ratio has been chosen because psychological tests have shown that it best matches the human visual field. It also enables use of existing cinema film formats as additional source material, since this is the same aspect ratio used in normal 35 mm film. Figure 16.6(a) shows how the aspect ratio of HDTV compares with that of conventional television, using the same resolution, or the same surface area as the comparison metric.

To achieve the improved resolution the video image used in HDTV must contain over 1000 lines, as opposed to the 525 and 625 provided by the existing NTSC and PAL systems. This gives a much improved vertical resolution. The exact value is chosen to be a simple multiple of one or both of the vertical resolutions used in conventional TV. However, due to the higher scan rates the bandwidth requirement for analogue HDTV is approximately 12 MHz, compared to the nominal 6 MHz of conventional TV, Table 1.2.

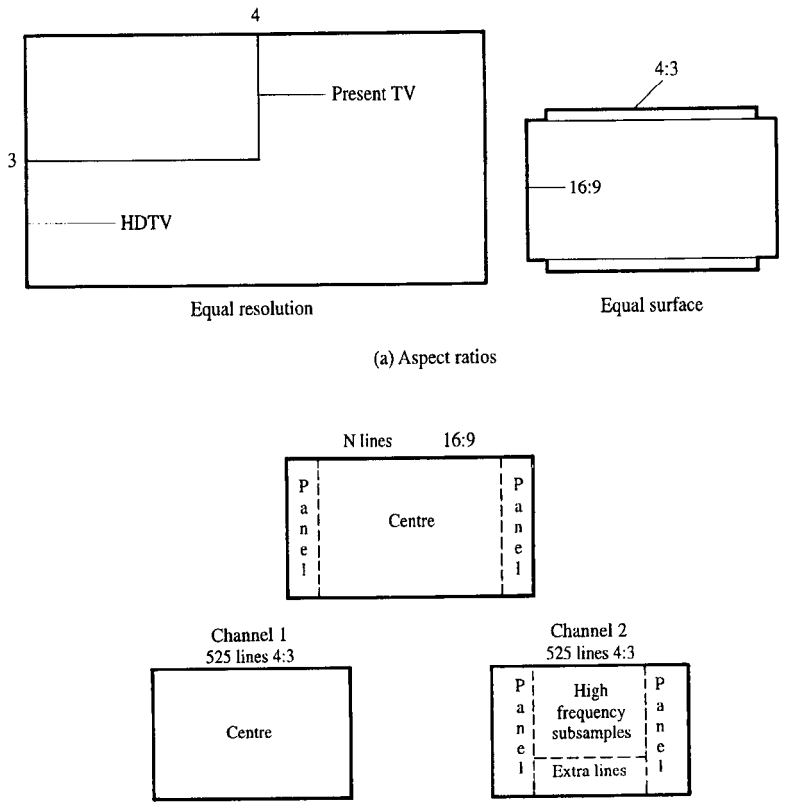

Figure 16.6 *(a) Comparison between conventional TV and HDTV, (b) 2-channel transmission.*

The introduction of a non-compatible TV transmission format for HDTV would require the viewer either to buy a new receiver, or to buy a converter to receive the picture on their old set. The initial thrust in Japan was towards an HDTV format which is compatible with conventional TV standards, and which can be received by conventional receivers, with conventional quality. However, to get the full benefit of HDTV, a new wide screen, high resolution receiver has to be purchased.

One of the principal reasons that HDTV is not already common is that a general standard has not yet been agreed. The XVIth CCIR plenary assembly recommended the adoption of a single, worldwide standard for high definition television. Unfortunately, Japan, Europe and North America are all investing significant time and money in their own systems based on their own, current, conventional TV standards and other national considerations.

16.4.2 Studio standards

Initially there were two main proposals for a worldwide HDTV *studio* system, with characteristics as shown in Table 16.1. The Japanese broadcasting company, NHK, has proposed one of the systems while the joint European project, Eureka, has proposed an alternative standard.

Table 16.1 *Proposed HDTV studio production standards.*

	Europe	*North America, Japan*
Total lines/picture	1250	1125
Active lines/picture	1192	1035
Scanning method	1:1	2:1 Interlaced
Aspect ratio	16:9	16:9
Field frequency (Hz)	50	60
Line frequency (kHz)	62.5	33.75
Samples/active line	1920	1920

The European standard uses a 50 Hz field rate to provide relatively easy conversion to both 60 and 50 Hz conventional, and HDTV, systems. It is also well suited to transfer from film. 1250 lines were chosen as this is exactly double the number of lines of the European conventional standards. A conversion to the American 525 lines standard is slightly more difficult, involving a ratio of 50/21.

16.4.3 Transmissions

In order to achieve compatibility with conventional TV it is proposed to split the HDTV information and transmit it in two separate channels. When decomposing a 1250 line studio standard, for US transmission, the line interpolator would extract the centre portion of every second line of the top 1050 lines, and send it in channel 1 as the reduced resolution 525 line image, left part of Figure 16.6(b). The panel extractor then removes

the side panels, which make up the extra width of the picture, adds in the bottom 200 lines which were not sent with channel 1 plus the missing alternate lines and sends this as channel 2. These two channels are then reconstructed in the TV receiver. An added advantage of a two channel system is that it would allow more than one picture to be displayed simultaneously.

Recent European developments have swung in favour of a digital HDTV transmission standard. This will most likely use the compression schemes described in section 16.7 combined with QAM or QPSK for cable or satellite systems [Forrest] or simultaneous orthogonal frequency division multiplex (OFDM) transmission [Aldard and Lassalle] for the more congested terrestial systems. This replaces one high bit rate signal with many (1,000 to 8,000) parallel, low bit rate, channels using orthogonal carriers. OFDM can better handle multipath fades as these only degrade a fraction of the parallel channels at any one time. With convolutional coded FECC transmissions, errors can be corrected subsequently in the receiver.

More recently the US grand alliance has pooled all interests in US HDTV into a single consortium to develop the best possible US HDTV system. This will use digital video compression in line with the MPEG standard (see later section 16.7.3) to compress the digital video into a 20 Mbit/s signal which will be transmitted in the conventional 6 MHz wide channel using high level modulation schemes such as those described in Chapter 11. The idea here is to use currently unallocated TV channels with a simultaneously broadcast analogue system rather than augment the existing NTSC system.

16.5 Digital video

Video in either RGB or YUV format can be produced in digital form [Forrest]. In this case discrete samples of the analogue video signal are digitised, to give a series of PCM words which represent the pixels. The words are divided into three fields representing each of the three signals RGB or YUV, section 16.2. Pixel word sizes range from 8 bits to 24 bits. Typical configurations include:

24 bits – where R = G = B = 8 bits and
16 bits – where Y = 8 bits and U = V = 4 bits.

In the latter case the chrominance signals are transmitted with less accuracy than the luminance but the eye is not able to discern the degradation. It is then possible either to store the samples in a memory device (e.g. compact disc (CD)), or to transmit them as a digital signal. Figure 16.7 shows the entropy (see section 9.2.4) of the luminance and chrominance signals in a video sequence comprising a fast moving sports sequence. Cb is the chrominance signal B – Y and Cr is R – Y. Scenes with less movement, such as a head and shoulders newsreader, have correspondingly lower entropy values which can be taken full advantage of if frame to frame differential coding is employed. Note, in Figure 16.7, that the luminance information (Y) requires a higher accuracy in the quantisation operation than does the chrominance information.

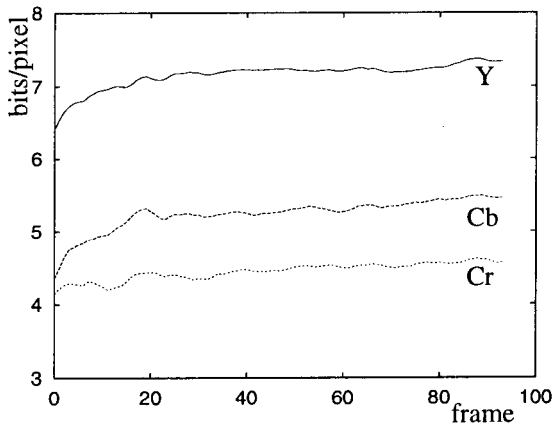

Figure 16.7 *Typical entropy measures in a colour video image sequence.*

It would initially seem ideal to use a computer network, or computer, modem and telephone network for video delivery. Unfortunately, the amount of data which would need to be transmitted is usually excessive. For a conventional TV system, the equivalent digital bit rate is around 140 Mbit/s which is not compatible with a standard modem. Even if the quality of the received image is reduced, data rates are still high. For example, if an image with a reduced resolution of 256×256 pixels is considered where each pixel consists of 16 bits (YUV) and a standard video frame rate of 25 frames per second is used, the bit rate R_b is given by:

$$R_b = 256 \times 256 \times 16 \times 25 = 26.2 \text{ Mbit/s} \tag{16.5}$$

Even a single frame requires 1 Mbit, or 132 kbyte of storage. For full resolution (720 pixel by 480 pixel) ITU-R 601 PAL TV with 8-bit quantisation for each of the three colour components, the rate grows to 207 Mbit/s. A further problem exists in storage and display as the access rate of CD-ROM is 120 kbyte/s while fast hard disc is only 500 kbyte/s. Fortunately, a general solution to the problem of data volume exists. In our sports sequence example, 16 bits were needed to represent each YUV pixel of the image (Figure 16.7). However, the long term average entropy of each pixel is considerably smaller than this and so data compression can be applied.

16.6 Video data compression

Data rate reduction is achieved by exploiting the *redundancy* of natural image sequences which arises from the fact that much of a frame is constant or predictable, because most of the time changes *between* frames are small. This is demonstrated in Figure 16.8 which shows a sequence of frames with a uniform background. The pixels representing the background are identical. The top sequence of frames is very similar in many respects, expect for some movement of the subject. The lower sequence has more movement

which increases the entropy. It is thus inefficient to code every image frame into 1 Mbit of data and ignore the predictive properties possessed by the previous frames. Also there are few transmission channels for which we can obtain the 25 to 200 MHz bandwidth needed for uncompressed, digital, TV. For example the ISDN digital telephone channel has 128 kbit/s capacity, while ISDN primary access rate is 2 Mbit/s and the full STM-1 rate is only 45 Mbit/s (see Chapter 19).

Removal of redundancy is achieved by image sequence coding. Compression operations (or algorithms) may work within a single frame (intra-frame), or between frames of a sequence (inter-frame), or a combination of the two. Practical compression systems tend to be hybrid in that they combine a number of different compression mechanisms. For example, the output of an image compression algorithm may be Huffman coded (Chapter 9) to further reduce the final output data rate.

We will look firstly at a few of the basic compression principles, before progressing to review practical systems. General coding techniques such as DPCM discussed in sections 5.8.2 and 5.8.3 are also applicable to image compression. Figure 16.9 shows the interframe difference between two images. DPCM would achieve 2 to 3 times compression compared to conventional PCM quantisation. However, video compression requires 20:1 to 200:1 compression ratios. Video compression and expansion equipment is often referred to as a 'video CODEC' inferring the ability to both transmit and receive images. In practice, not all equipment is able to do this, and the term is sometimes used to refer to the transmitter (coder) or receiver (decoder) only.

16.6.1 Run-length coding

This intra-frame compression algorithm is best suited to graphic images, or video images with large sections made up of identical pixels. The algorithm simply detects the presence of a sequence of identical pixel values (usually operating on the luminance component only), and notes the start point, and number of pixels in the run (the 'run length'). This information is then transmitted in place of the original pixel values. (The

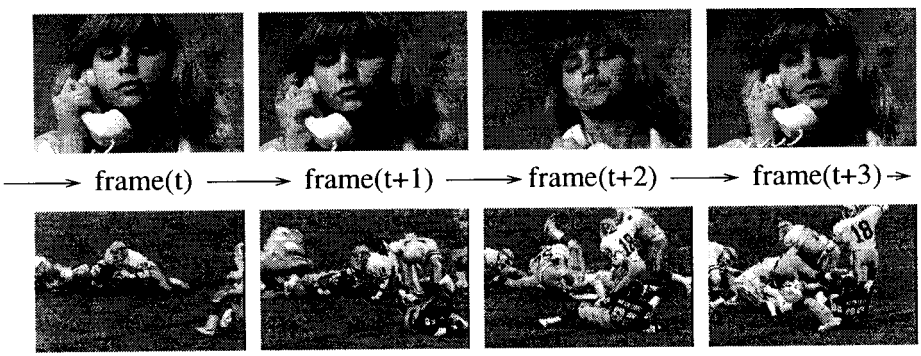

Figure 16.8 *Redundancy in video image sequence.*

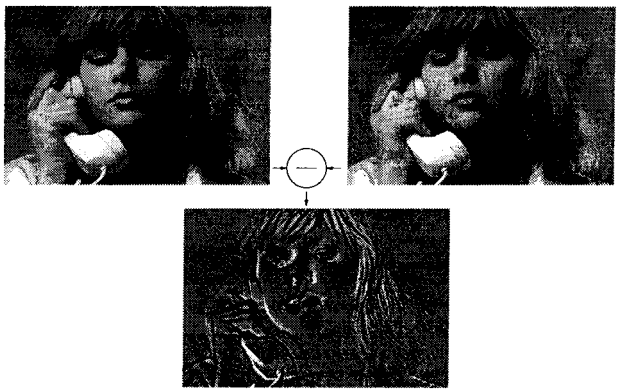

Figure 16.9 *Difference image between two frames in a video image sequence.*

technique was described as applied in facsimile transmission in Chapter 9.) To achieve a significant coding gain the run lengths must be large enough to provide a saving when considering the additional overhead of the addressing and control information which is required.

16.6.2 Conditional replenishment

This is an intra-frame coding algorithm which requires a reference frame to be held at the transmitter. The algorithm first divides the frame into small elements, called blocks, although lines can be used. Each pixel element is compared with the same location in the reference frame, and some measure of the difference between the pixel elements is calculated. If this is greater than a decision threshold, the pixel is deemed to have changed and the new value is sent to the receiver and updated in the reference frame. If the difference is not greater than the threshold, no data is sent. The frame displayed is therefore a mixture of old and new pixel element values.

16.6.3 Transform coding

As its name suggests, transform coding attempts to convert or transform the input samples of the image from one domain to another [Clarke 1985]. It is usual to apply a two-dimensional (2-D) transform to the two dimensional image and then quantise the transformed output samples. Note that the transform operation does not in itself provide compression; often one transform coefficient is output for each video amplitude sample input, as in the Fourier transform (Chapters 2 and 13). However, many grey scale patterns in a two dimensional image transform into a much smaller number of output samples which have *significant* magnitude. (Those output samples with insignificant magnitude need not be retained.) Furthermore, the quantisation resolution of a particular output sample which *has* significant magnitude can be chosen to reflect the importance of

that sample to the overall image.

To reconstruct the image, the coefficients are input to an inverse quantiser and then inverse transformed such that the original video samples are reconstructed, Figure 16.10. Much work has been undertaken to discover the optimum transforms for these operations, and the Karhunen-Loeve transform [Clarke 1985] has been identified as one which minimises the overall mean square error between the original and reconstructed images. Unfortunately, this is complex to implement in practice and alternative, sub-optimum transforms are normally used.

These alternative transforms, which include sine, cosine and Fourier transforms amongst others, must still possess the property of translating the input sample energy into another domain. The cosine transform has been shown to produce image qualities similar to the Karhunen-Loeve transform for practical images (where there is a high degree of interframe pixel correlation) and is now specified in many standardised compression systems.

The 2-D discrete cosine transform (DCT) [Clarke 1985] can be defined as a matrix with elements:

$$C_{nk} = \begin{cases} \sqrt{\frac{1}{N}} \cos\left[\frac{n(2k+1)\pi}{2N}\right], & n = 0, \quad 0 \le k \le N-1 \\ \sqrt{\frac{2}{N}} \cos\left[\frac{n(2k+1)\pi}{2N}\right], & 1 \le n \le N-1, 0 \le k \le N-1 \end{cases} \tag{16.6}$$

where N is the number of samples in one of the dimensions of the normally square transform data block. The fact that the video data comprises real (pixel intensity) rather than complex sample values favours the use of the DCT over the DFT.

The next stage in a transform based video compression algorithm is to quantise the output frequency samples. It is at this stage that compression can occur. Recall that the output samples now represent the spatial frequency components of the input 'signal' [Clarke, 1985, Grant *et al.*]. The first output sample represents the DC, or average, value of the 'signal' and is referred to as the DC coefficient. Subsequent samples are AC coefficients of increasing spatial frequency. Low spatial frequency components imply slowly changing features in the image while sharp edges or boundaries demand high spatial frequencies.

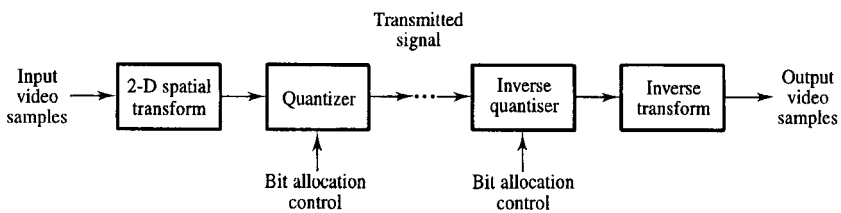

Figure 16.10 *Simplified block diagram illustrating transform coder operation.*

Much of the image information registered by the human visual process is contained in the lower frequency coefficients, with the DC coefficient being the most prominent. The quantisation process therefore involves allocating a different number of bits to each coefficient. The DC coefficient is allocated the largest number of bits (i.e. it is given the highest resolution), with increasingly fewer bits being allocated to the AC coefficients representing increasing frequencies. Indeed, many of the higher frequency coefficients are allocated 0 bits (i.e. they are simply not required to reconstruct an acceptable quality image).

The combined operations of transformation, quantisation and their inverses are shown in Figure 16.10. Note that the quantisation and inverse quantisation operations may be adaptive in the number of bits allocated to each of the transformed values. Modification of the bit allocations allows the coder to transmit over a channel of increased or decreased capacity as required, perhaps due to changes in error rates or a change of application. With the advent of ATM (section 19.7.2), such a variable bit rate (bursty) signal can now be accommodated.

Practical transform coders include other elements in addition to the transformation and quantisation operations described above. Further compression can be obtained by using variable length (e.g. Huffman) type coding operating on the quantiser output. In addition, where coding systems operate on continuous image sequences, the transformation may be preceded by inter-frame redundancy removal (e.g. movement detection) ensuring that only changed areas of the new frame are coded. Some of these supplimentary techniques are utilised in the practical standard compression mechanisms described below.

16.7 Compression standards

16.7.1 COST 211

This CODEC specification was developed in the UK by British Telecom in conjunction with GEC, and the CODEC has been used extensively for professional video conference applications. Its output bit rate is high compared with more recent developments based on transform techniques. A communication channel of at least 340 kbit/s is required for the COST 211 CODECs, the preferred operation being at the 2 Mbit/s ISDN primary access rate (Chapter 19). The COST 211 CODEC provides full frame rate (25 frame/s) video assuming a conference scene without excessive changes between frames. The CODEC has a resolution of 286 lines, each of 255 pixels. The compression uses intra-frame algorithms, initially utilising conditional replenishment where the changed or altered pixel information is transmitted, followed by DPCM coding (section 5.8.2) which is applied to the transmitted pixels. Finally, the output is Huffman encoded.

Further compression modes are available, including sub-sampling (skipping over sections of the frame, or even missing complete frames) to enable the CODEC output to be maintained at a constant rate if replenishment threshold adjustment is not sufficient.

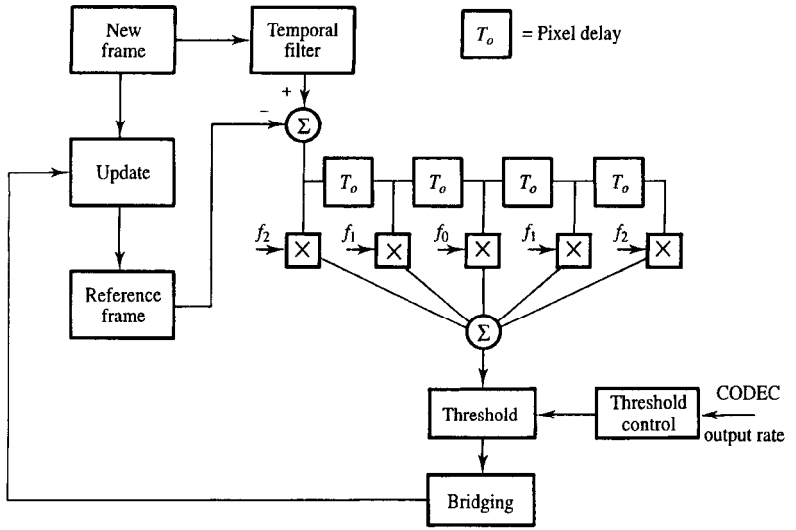

Figure 16.11 *COST 211 codec movement detector.*

The conditional replenishment, or movement detector, part of the CODEC is shown in Figure 16.11. This operates on a pixel-by-pixel basis and calculates the weighted sum of 5 consecutive pixel differences. The difference sum is then compared with a variable threshold value to determine if the pixel has changed. The threshold is adjusted to keep the CODEC output rate constant. Changed pixels are both transmitted, and stored in the local reference frame at the transmitter.

16.7.2 JPEG

JPEG is an international standard for the compression and expansion of single frame monochrome and colour images. It was developed by the Joint Photographic Experts Group (JPEG) and is actually a set of general purpose techniques which can be selected to fulfil a wide range of different requirements. However, a common core to all modes of operation, known as the baseline system, is included within JPEG. JPEG is a transform based coder and some, but not all, operating models use the DCT applied to blocks of 8×8 image pixels yielding 64 output coefficients. Figure 16.12 shows a 352-by-240 pixel image, partitioned into 330 macroblocks. The 16×16 pixel macroblocks are processed slice-by-slice in the DCT coder.

The coefficients are then quantised by a user defined quantisation table which specifies a quantiser step size for each coefficient in the range 1 to 255. The DC coefficient is coded as a difference value from that in the previous block, and the sequence of AC coefficients is reordered as shown in Figure 16.13(a) by zig-zag scanning in ascending order of spatial frequency, and hence decreasing magnitude, progressing from the more significant to the least significant components. Thus the quantiser step

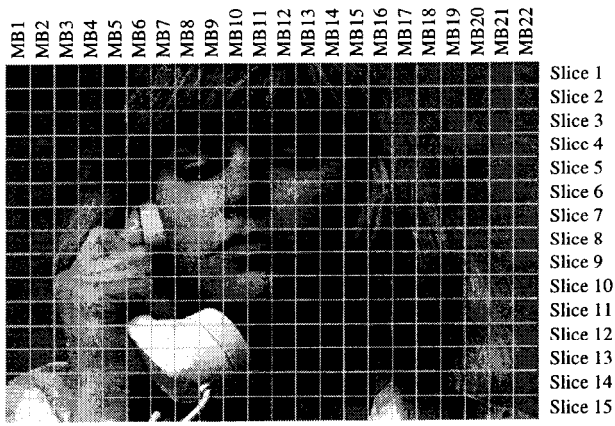

Figure 16.12 *Image partitioning into macroblocks (MB).*

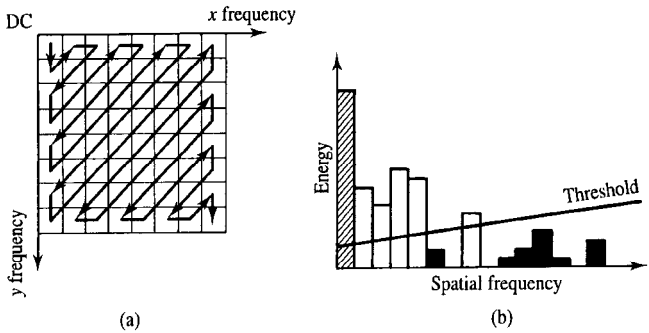

Figure 16.13 *(a) Zig-zag scanning to reorder transformed data into increasing spatial frequency components (b) and permit variable length coding of the scanned components.*

size increases and becomes coarser with increasing spatial frequency in the image to aid the final data compression stage, Figure 16.13(b).

The final stage consists of encoding the quantised coefficients according to their statistical probabilities or entropy (Chapter 9 and Figure 16.7). Huffman encoding is used in the baseline JPEG systems. JPEG also uses a predictor, measuring the values from three adjacent transformed pixels, to estimate the value of the pixel which is about to be encoded. The predicted pixel value is subtracted from the actual value and the difference signal is sent to the Huffman coder as in DPCM (Chapter 5).

JPEG compression ratios range from 2:1 to 20:1. Low compression ratios achieve lossless DPCM coding where the reconstructed image is indistinguishable from the original while high compression ratios can reduce the storage requirements to only ¼ bit per pixel and the transmission rate, for image sequences, to 2 Mbit/s. VLSI JPEG chips

existed in 1993 which could process data at 8 Mbyte/s to handle 352×288 pixel (¼ full TV frame) images at 30 frame/s. As JPEG is a single frame coder it is not optimised to exploit the interframe correlation in image sequence coding. The following schemes are therefore preferred for compression of video data.

16.7.3 MPEG

This video coding specification was developed by the Motion Picture Experts Group (MPEG) as a standard for coding image *sequences* to a bit rate of about 1.5 Mbit/s for MPEG1 and 10 Mbit/s for MPEG2. The lower rate was developed, initially, for 352×288 pixel images because it is compatible with digital storage devices such as hard disc drives, compact discs, digital audio tapes, etc. The algorithm is deliberately flexible in operation, allowing different image resolutions, compression ratios and bit rates to be achieved. The basic blocks which can be used include:

- Motion compensation
- DCT
- Variable length coding

A simplified diagram of the encoding operation is shown in Figure 16.14. Because the algorithm is intended primarily for the storage of image *sequences* it incorporates motion compensation which is not included in JPEG. A compromise still exists between the need for high compression, and easy regeneration of randomly selected frame sequences. MPEG therefore allows frames to be coded in one of three ways, Figure 16.15.

Intra-frames (I-frames), which are coded independently of other frames, allow random access, but provide limited compression. They form the start points for replay sequences. Unidirectional predictive coded frames (P-frames) can achieve motion prediction from *previous* reference frames and hence, with the addition of motion compensation, the bit rate can be reduced. Bidirectionally predictive coded frames (B-frames) provide greatest compression but require two reference frames, one previous frame and one future frame in order to be regenerated. Figure 16.15 shows such an I, P, B picture sequence. The precise combination of I, P and B frames which are used depends upon the specific application.

The motion prediction and compensation uses a block based approach on 16×16 pixel block sizes as in Figure 16.12. The concept underlying picture motion

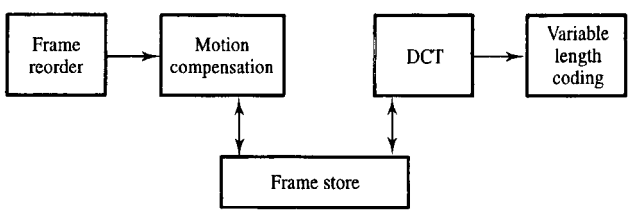

Figure 16.14 *Block diagram for simplified MPEG encoder.*

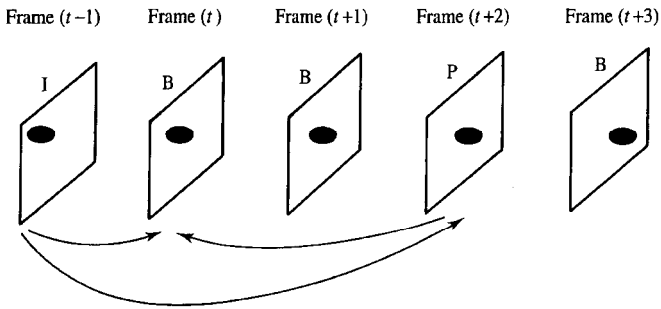

Figure 16.15 *I, P, B image frames, as used in the MPEG coder.*

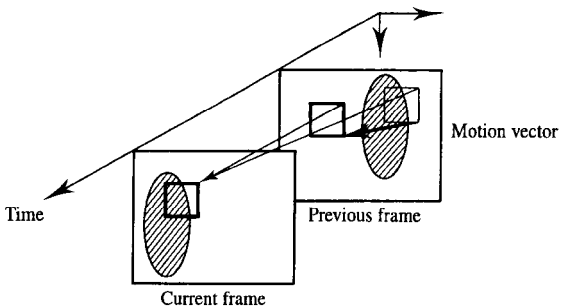

Figure 16.16 *Estimation of the motion vector, between consecutive frames, with object movement.*

compensation is to estimate the motion vector, Figure 16.16, and then use this to enable the information in the previous frame to be used in reconstructing the current frame, minimising the need for new picture information. MPEG is a continuously evolving standard. MPEG2 is aimed at HDTV where partial images are combined into the final image. Here the processing is dependent on the image content to enhance the compression capability and permit use with ATM at a P_b of 10^{-4}. MPEG2 data rates vary between VHS video cassette player quality at 1.5 Mbit/s through to 30 Mbit/s for HDTV images. It is generally recognised that fast moving sports scenes require the higher bit rates of 6 to 8 Mbit/s and, by reducing these to 1.5 Mbit/s, then the quality degrades, to much like that of a VHS video cassette player. MPEG3 is included within MPEG2. MPEG4 is now addressing very low bit rates for wireless video on the PSTN and should be standardised around 1998.

16.7.4 H.261 and H.263

The H.261 algorithm was developed for the purpose of image transmission rather than image storage. It is designed to produce a constant output of $p \times 64$ kbit/s, where p is an integer in the range 1 to 30. This allows transmission over a digital network or data link

of varying capacity. It also allows transmission over a single 64 kbit/s digital telephone channel for low quality videotelephony, or at higher bit rates for improved picture quality. The basic coding algorithm is similar to that of MPEG in that it is a hybrid of motion compensation, DCT and straightforward DPCM (intra-frame coding mode), Figure 16.17, without the MPEG I, P, B frames. The DCT operation is performed at a low level on 8 × 8 blocks of error samples from the predicted luminance pixel values, with sub-sampled blocks of chrominance data. The motion compensation is performed on the macroblocks of Figure 16.12, comprising four of the previous luminance and two of the previous chrominance blocks.

H.261 is widely used on 176 × 144 pixel images. The ability to select a range of output rates for the algorithm allows it to be used in different applications. Low output rates (p = 1 or 2) are only suitable for face-to-face (videophone) communication. H.261 is thus the standard used in many commercial videophone systems such as the UK BT/Marconi Relate 2000 and the US ATT 2500 products. Video-conferencing would require a greater output data rate ($p > 6$) and might go as high as 2 Mbit/s for high quality transmission with larger image sizes.

A further development of H.261 is H.263 for lower fixed transmission rates. This deploys arithmetic coding in place of the variable length coding in Figure 16.17 and, with other modifications, the data rate is reduced to only 20 kbit/s.

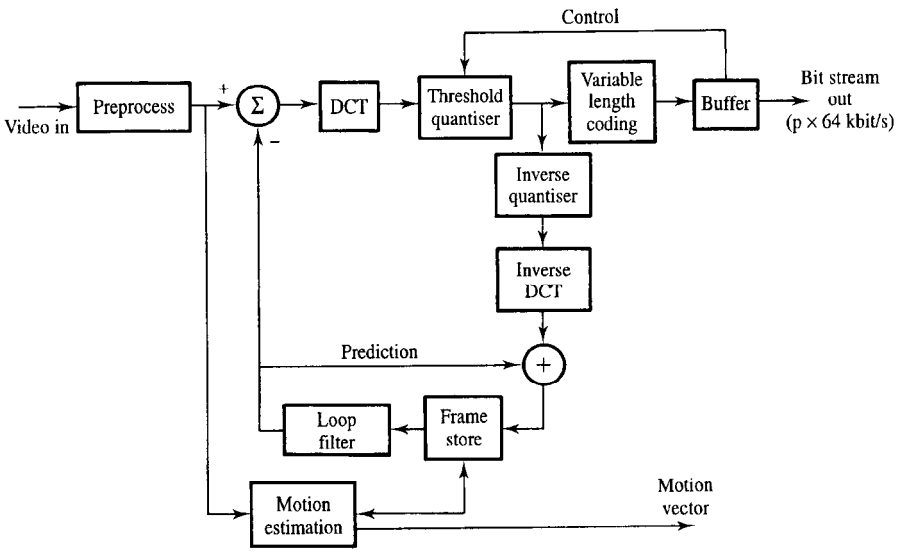

Figure 16.17 *Schematic representation of the H.261 encoder standard.*

16.7.5 Model based coding

At the very low bit rates (20 kbit/s or less) associated with video telephony, the requirements for image transmission stretch the compression techniques described earlier to their limits. In order to achieve the necessary degree of compression they often require reduction in spatial resolution or even the elimination of frames from the sequence. Model based coding (MBC) attempts to exploit a greater degree of redundancy in images than current techniques, in order to achieve significant image compression but without adversely degrading the image content information. It relies upon the fact that the image quality is largely subjective. Providing that the appearance of scenes within an observed image is kept at a visually acceptable level, it may not matter that the observed image is not a precise reproduction of reality.

One MBC method for producing an artificial image of a head sequence utilises a feature codebook where a range of facial expressions, sufficient to create an animation, are generated from sub-images or templates which are joined together to form a complete face. The most important areas of a face, for conveying an expression, are the eyes and mouth, hence the objective is to create an image in which the movement of the eyes and mouth is a convincing approximation to the movements of the original subject. When forming the synthetic image, the feature template vectors which form the closest match to those of the original moving sequence are selected from the codebook and then transmitted as low bit rate coded addresses.

By using only 10 eye and 10 mouth templates, for instance, a total of 100 combinations exists implying that only a 6-bit codebook address need be transmitted. It has been found that there are only 13 visually distinct mouth shapes for vowel and consonant formation during speech. However, the number of mouth sub-images is usually increased, to include intermediate expressions and hence avoid step changes in the image.

Another common way of representing objects in three-dimensional computer graphics is by a net of interconnecting polygons. A model is stored as a set of linked arrays which specify the coordinates of each polygon vertex, with the lines connecting the vertices together forming each side of a polygon. To make realistic models, the polygon net can be shaded to reflect the presence of light sources.

The wire-frame model [Welch 1991] can be modified to fit the shape of a person's head and shoulders. The wire-frame, composed of over 100 interconnecting triangles, can produce subjectively acceptable synthetic images, providing that the frame is not rotated by more than 30° from the full-face position. The model, shown in Figure 16.18, uses smaller triangles in areas associated with high degrees of curvature where significant movement is required. Large flat areas, such as the forehead, contain fewer triangles. A second wire-frame is used to model the mouth interior.

A synthetic image is created by texture mapping detail from an initial full-face source image, over the wire-frame. Facial movement can be achieved by manipulation of the vertices of the wire-frame. Head rotation requires the use of simple matrix operations upon the coordinate array. Facial expression requires the manipulation of the features controlling the vertices.

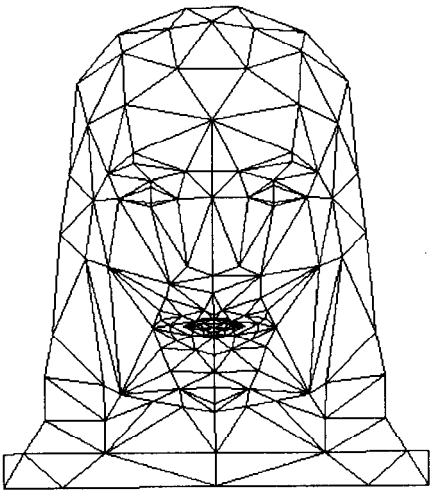

Figure 16.18 *Image representation by wire frame model.*

This model based feature codebook approach suffers from the drawback of codebook formation. This has to be done off-line and, consequently, the image is required to be prerecorded, with a consequent delay. However, the actual image sequence can be sent at a very low data rate. For a codebook with 128 entries where 7 bits are required to code each mouth, a 25 frame/s sequence requires less than 200 bit/s to code the mouth movements. When it is finally implemented, rates as low as 1 kbit/s are confidently expected from MBC systems, but they can only transmit image sequences which match the stored model, e.g. head and shoulders displays.

16.8 Packet video

Packet video is the term used to describe low-bit-rate encoded video, from one of the above CODECs, which is then transmitted over the ISDN as packet data similar to the packet speech example shown later in Figure 17.14. Due to the operation of the CODEC and the nature of the video signal, the resulting data rate varies considerably between low data rates when the image sequence contains similar frames, as in Figure 16.8, to high data rates when there is a sudden change in the sequence and almost the entire image frame must be encoded and transmitted. This is called variable bit rate (VBR) traffic, Figure 16.19, which is accommodated by packet transmission on an ATM network, Chapter 19. Figure 16.19 shows the regular variation in bit rate in the MPEG coder arising from quantising the I, B, P frames, Figure 16.15, plus the major alterations in rate due to motion or dramatic scene changes in the video sequence. At 30 frame/s this represents an overall variation in the required transmission rate between 600 kbit/s and 6 Mbit/s.

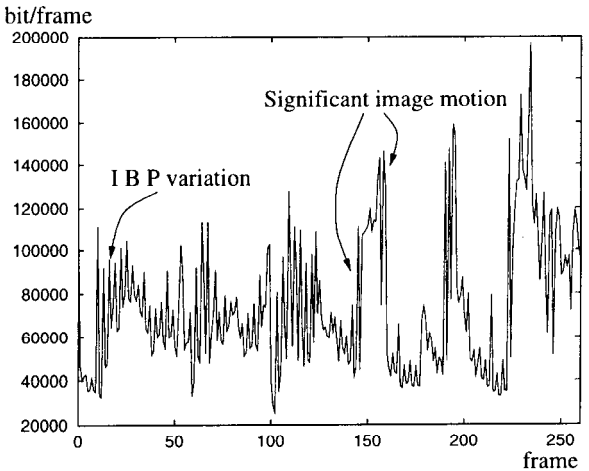

Figure 16.19 *Variable rate bursty traffic from a colour video image codec. Note the regular short-term variation in the MPEG coded data and the major scene changes.*

VBR systems rely on the fact that their average transmission rate has a modest bandwidth requirement. Hence if many VBR sources are statistically multiplexed then their individual bit rate variations will be averaged and they can be carried over a system with a transmission bandwidth which is compatible with the modest *average* bandwidth requirement of each source. Tariffs will then be levied which use this modest data rate requirement, rather than providing each user with access to a peak data rate channel and incurring correspondingly high costs.

The principles underlying packet video are similar to those described in Chapter 17. We must inevitably accept some loss of video data when the overall demand for transmission bandwidth exceeds the available channel allocation since arbitrarily large amounts of data cannot be held within the finite length of the buffer in Figure 16.17 (and Figure 17.1). Much current research is aimed at investigating this problem as the ISDN system expands and ATM techniques start to become more widely used. Some of this work is aimed at the design of the video encoding algorithm to control the statistics of the output VBR traffic; other work is aimed at investigating coding methods which split the video data into high and low priority traffic.

The overall protocol [Falconer and Adams], which corresponds to the lowest two layers of the OSI reference model (see section 18.4.1), represents one potential method of achieving different priority of traffic. By careful control of the video encoding technique we can ensure that high priority bits are allocated to the major video image features while low priority bits add more detail to this basic image. Techniques such as this allow speech and video data to be transmitted, with some packet loss, while still providing a quality of service which is acceptable to the user, for a transmission cost which is much lower than that of providing the full bandwidth requirement for 100% of the time.

16.9 Summary

The smallest elemental area of a two dimensional image which can take on characteristics independent of other areas is called a picture cell or pixel. The fundamental requirement of an image transmission system is the ability to convey a description of each pixel's characteristics: intensity, colour and location in space. For video images, information giving the pixel's location in time is also required.

A pixel from a colour image may be represented by three colour intensity signals (red, green and blue) or by a white (red + green + blue) intensity signal (luminance) and two colour difference (chrominance) signals (white – red and white – blue). The colour difference signals may be transformed geometrically, using the colour triangle, into saturation and hue. The luminance plus chrominance representation makes colour image transmission compatible with monochrome receivers and is therefore used for conventional TV broadcasts.

In the UK the two chrominance signals of a TV broadcast are transmitted using quadrature double sideband suppressed carrier modulation of a sub-carrier. Since this requires a phase coherent reference signal for demodulation, the reconstructed image is potentially susceptible to colour errors caused by phase distortion, within the chrominance signal frequency band, arising from multipath propagation. These errors can be reduced by reversing the carrier phase of one of the quadrature modulated chrominance signals on alternate picture lines which gives the system its name – 'phase alternate line' or PAL. The PAL signal transmits 25 frames per second, each frame being made up of 625 lines. The frame is divided into two fields, one containing even lines only and the other containing odd lines only. The field rate is therefore 50 Hz. The total bandwidth of a PAL picture signal is approximately 6 MHz.

HDTV may be either analogue (for compatibility with traditional broadcast receivers), digital or mixed. The picture resolution of HDTV is about twice that of conventional TV as is the required (analogue) signal bandwidth. Compatibility with conventional TV receivers can be maintained but it is much more likely that HDTV will ultimately adopt an all-digital transmission system.

Digital video can be implemented by sampling and digitising an analogue video signal. Typically, the luminance signal will be subject to a finer quantisation process than the chrominance signals (e.g. 8 bit versus 4 bit quantisation). Unprocessed digital video signals have a bandwidth which is too large for general purpose telecommunications transmissions channels. To utilise digital video for services such as videophone or teleconferencing the digital signal therefore requires data compression. Several coding techniques for redundancy removal, including run-length coding, conditional replenishment, transform coding, Huffman coding and DPCM may be used, either in isolation or in combination, to achieve this.

Several international standards exist for the transmission and/or storage of images. These include JPEG which is principally for single images (i.e. stills) and MPEG which is principally for moving pictures (i.e. video).

Very low bit rate transmission of video pictures can be achieved, using aggressive coding, if only modest image reproduction quality is required. Thus the BT/Marconi

standard H.261 CODEC allows videophone services to be provided using a single 64 kbit/s digital telephone channel. Even lower bit rate video services can be provided using model based coding schemes. These are analogous to the low bit rate speech vocoder techniques described in Chapter 9.

Variable bit rate transmissions, which may arise as a result of video coding, can be accommodated by packet transmission over ATM networks. Predicting the overall performance of packet systems requires queuing theory which is discussed in Chapter 17. Networks, and the foundations of ATM data transmission, are discussed in Chapters 18 and 19 respectively.

16.10 Problems

16.1. A 625-line black and white television picture may be considered to be composed of 550 picture elements (pixels) per line. (Assume that each pixel is equiprobable among 64 distinguishable brightness levels.) If this is to be transmitted by raster scanning at a 25 Hz frame rate calculate, using the Shannon-Hartley theorem of Chapter 11, the minimum bandwidth required to transmit the video signal, assuming a 35 dB signal to noise ratio on reception. [4.44 MHz]

16.2. What is the minimum time required to transmit one of the picture frames described in Problem 16.1 over a standard 3.2 kHz bandwidth telephone channel, as defined in Figure 5.12, with 30 dB SNR? [65 s]

16.3. A colour screen for a US computer aided design product may be considered to be composed of a 1,000 × 1,000 pixel array. Assume that each pixel is coded with straightforward 24-bit colour information. If this is to be refreshed through a 64 kbit/s ISDN line. Calculate the time required to update the screen with a new picture. [6.25 min]

16.4. Derive an expression for the bandwidth or effective data rate for a television signal for the following scan parameters: frame rate = P frame/s; number of lines per frame = M; horizontal blanking time = B; horizontal resolution = x pixels/line where each pixel comprises one of k discrete levels. Hence calculate the bandwidth of a television system which employs 819 lines, a 50 Hz field rate, a horizontal blanking time of 18% of the line period and 100 levels. The horizontal resolution is 540 pixels/line. [75.5 Mbit/s]

Part Four

Networks

Part Four is devoted to communication networks which now exist on all scales from geographically small LANs to the global ISDN.

It starts with a discussion, in Chapter 17, of queuing theory which may be used to predict the delay suffered by digital information packets as they propagate through a data network.

Chapter 18 describes the topologies and protocols employed by networks to ensure the reliable, accurate and timely, delivery of information packets between network terminals. Rings, buses and their associated medium access protocols are discussed, and international standards such as ISO OSI, X.25 and FDDI are described. The optical transmission medium, which now forms an integral part of many communications networks, is also examined.

Part Four ends, in Chapter 19, by examining public networks. The current plesiochronous digital hierarchy (PDH) is reviewed before introducing a more detailed discussion of the new synchronous digital heirarchy (SDH) which will gradually come to replace it. The Chapter concludes with a brief discussion of PSTN/PDN data access techniques including the ISDN standard, ATM, and the probable future development of the local loop.