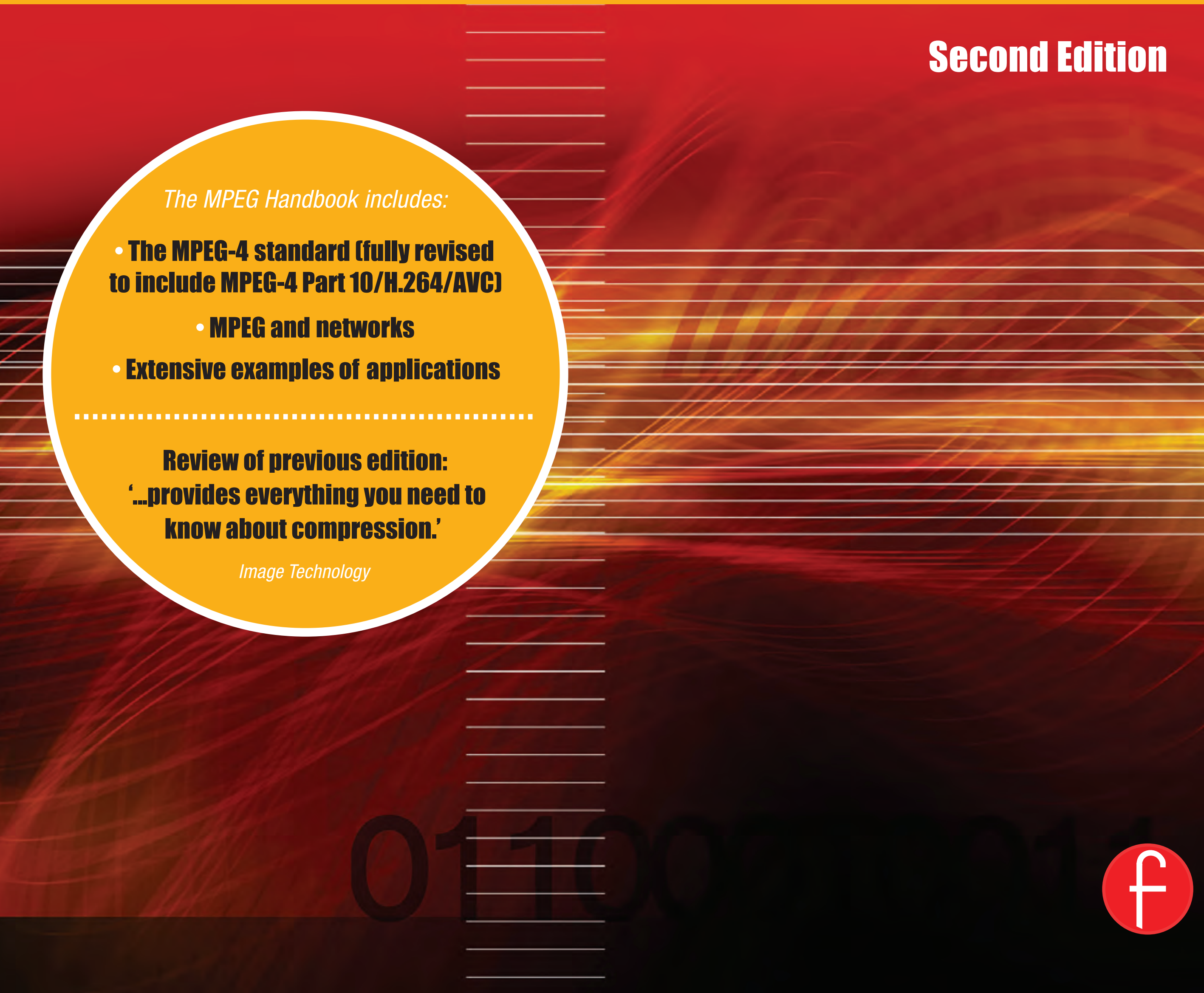
John Watkinson

THE MPEG HANDBOOK

MPEG-1 MPEG-2 MPEG-4 (MPEG-4 Part 10/H.264/AVC included)

Second Edition



The MPEG Handbook includes:

- **The MPEG-4 standard (fully revised to include MPEG-4 Part 10/H.264/AVC)**
- **MPEG and networks**
- **Extensive examples of applications**

Review of previous edition:
'...provides everything you need to know about compression.'
Image Technology

The MPEG Handbook

For Howard and Matthew

The MPEG Handbook

MPEG-1, MPEG-2, MPEG-4

Second edition

John Watkinson

Focal Press
Taylor & Francis Group
NEW YORK AND LONDON

First published 2001
Second edition 2004

This edition published 2013
by Focal Press
70 Blanchard Road, Suite 402, Burlington, MA 01803

Simultaneously published in the UK
by Focal Press
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Focal Press is an imprint of the Taylor & Francis Group, an informa business

Notices

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication Data
A catalogue record for this book is available from the Library of Congress

ISBN 13: 978-0-240-80578-8 (hbk)

Contents

This page □
intentionally
left blank

Preface

This book adds some material to the first edition which in turn completely revised the earlier book entitled *MPEG-2*. It is an interesting reflection on the rate at which this technology progresses that these frequent revisions are necessary. Perhaps after H.264 one could say that all the key work has been done and thereafter progress will be incremental. H.264, also known as AVC, is, of course, the impetus for this edition. The opportunity has also been taken to incorporate some improved explanations.

The approach of the book has not changed in the slightest. Compression is an arcane subject with its own library of specialist terminology that is generally accompanied by a substantial amount of mathematics. I have always argued that mathematics is only a form of shorthand, itself a compression technique! Mathematics describes but does not explain, whereas this book explains principles and then describes actual examples.

A chapter of fundamentals is included to make the main chapters easier to follow. Also included are some guidelines that have practically been found useful in getting the best out of compression systems. The reader who has endured this book will be in a good position to tackle the MPEG standards documents themselves, although these are not for the faint-hearted, especially the MPEG-4 and H.264 documents which are huge and impenetrable.

Acknowledgements

Information for this book has come from a number of sources to whom I am indebted. The publications of the ISO, AES and SMPTE provided essential reference material. Thanks also to the following for lengthy discussions and debates: Peter de With, Steve Lyman, Bruce Devlin, Mike Knee, Peter Kraniauskas and Tom MacMahon. The assistance of MicroSoft Corp. and Tektronix Inc. is also appreciated. Special thanks to Mikael Reichel.

1

Introduction to compression

1.1 What is MPEG?

MPEG is actually an acronym for the Moving Pictures Experts Group which was formed by the ISO (International Standards Organization) to set standards for audio and video compression and transmission.

Compression is summarized in Figure 1.1. It will be seen in (a) that the data rate is reduced at source by the *compressor*. The compressed data are then passed through a communication channel and returned to the original rate by the *expander*. The ratio between the source data rate and the channel data rate is called the *compression factor*. The term *coding gain* is also used. Sometimes a compressor and expander in series are referred to as a *compander*. The compressor may equally well be referred to as a *coder* and the expander a *decoder* in which case the tandem pair may be called a *codec*.

Where the encoder is more complex than the decoder, the system is said to be asymmetrical. Figure 1.1(b) shows that MPEG works in this way. The encoder needs to be algorithmic or adaptive whereas the decoder is 'dumb' and carries out fixed actions. This is advantageous in applications such as broadcasting where the number of expensive complex encoders is small but the number of simple inexpensive decoders is large. In point-to-point applications the advantage of asymmetrical coding is not so great.

The approach of the ISO to standardization in MPEG is novel because it is not the encoder which is standardized. Figure 1.2(a) shows that instead the way in which a decoder shall interpret the bitstream is defined. A decoder which can successfully interpret the bitstream is said to be *compliant*. Figure 1.2(b) shows that the advantage of standardizing the decoder is that over time encoding algorithms can improve yet compliant decoders will continue to function with them.

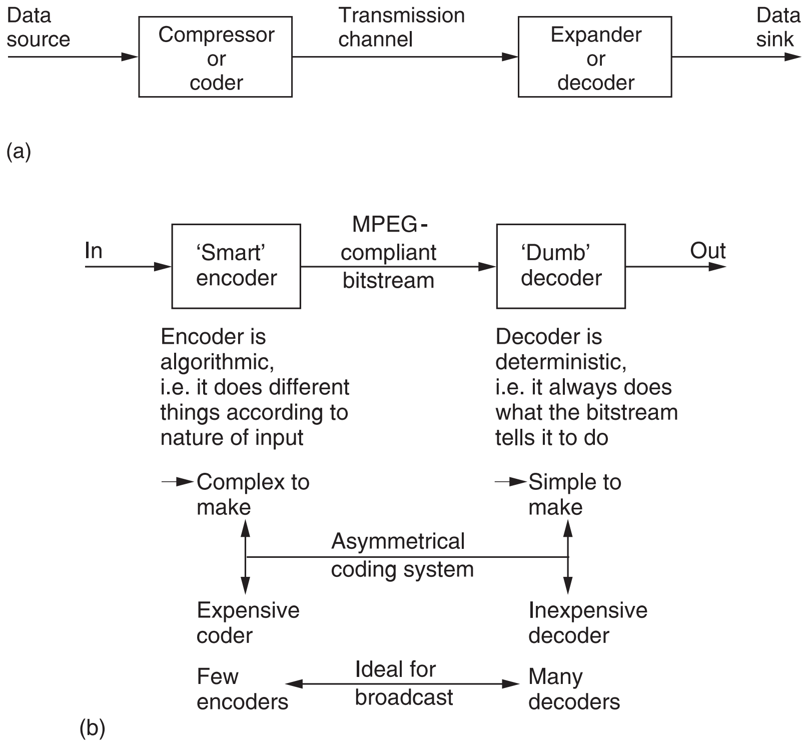


Figure 1.1 In (a) a compression system consists of compressor or coder, a transmission channel and a matching expander or decoder. The combination of coder and decoder is known as a codec. (b) MPEG is asymmetrical since the encoder is much more complex than the decoder.

It should be noted that a compliant decoder must correctly be able to interpret every allowable bitstream, whereas an encoder which produces a restricted subset of the possible codes can still be compliant.

The MPEG standards give very little information regarding the structure and operation of the encoder. Provided the bitstream is compliant, any coder construction will meet the standard, although some designs will give better picture quality than others. Encoder construction is not revealed in the bitstream and manufacturers can supply encoders using algorithms which are proprietary and their details do not need to be published. A useful result is that there can be competition between different encoder designs which means that better designs can evolve. The user will have greater choice because different levels of cost and complexity can exist in a range of coders yet a compliant decoder will operate with them all.

MPEG is, however, much more than a compression scheme as it also standardizes the protocol and syntax under which it is possible to combine or multiplex audio data with video data to produce a digital

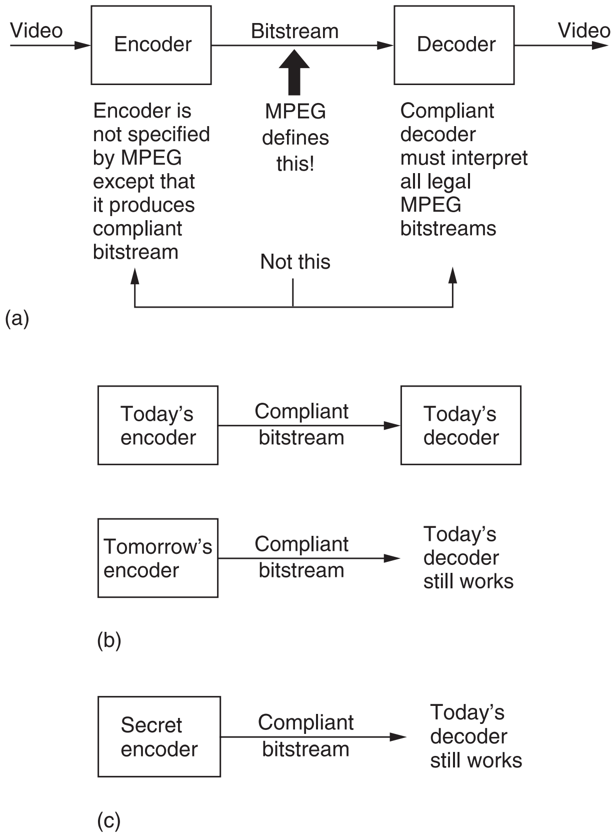

Video
Encoder
Bitstream
Decoder
Video
Encoder is not specified by MPEG except that it produces compliant bitstream
MPEG defines this!
Compliant decoder must interpret all legal MPEG bitstreams
Not this
(a)
Today's encoder
Compliant bitstream
Today's decoder
Tomorrow's encoder
Compliant bitstream
Today's decoder still works
(b)
Secret encoder
Compliant bitstream
Today's decoder still works
(c)

Figure 1.2 (a) MPEG defines the protocol of the bitstream between encoder and decoder. The decoder is defined by implication, the encoder is left very much to the designer. (b) This approach allows future encoders of better performance to remain compatible with existing decoders. (c) This approach also allows an encoder to produce a standard bitstream whilst its technical operation remains a commercial secret.

equivalent of a television program. Many such programs can be combined in a single multiplex and MPEG defines the way in which such multiplexes can be created and transported. The definitions include the metadata which decoders require to demultiplex correctly and which users will need to locate programs of interest.

As with all video systems there is a requirement for synchronizing or genlocking and this is particularly complex when a multiplex is assembled from many signals which are not necessarily synchronized to one another.

1.2 Why compression is necessary

Compression, bit rate reduction, data reduction and source coding are all terms which mean basically the same thing in this context. In essence the same (or nearly the same) information is carried using a smaller quantity or rate of data. It should be pointed out that in audio *compression* traditionally means a process in which the dynamic range of the sound is reduced. In the context of MPEG the same word means that the bit rate is reduced, ideally leaving the dynamics of the signal unchanged. Provided the context is clear, the two meanings can co-exist without a great deal of confusion.

There are several reasons why compression techniques are popular:

a Compression extends the playing time of a given storage device.
b Compression allows miniaturization. With fewer data to store, the same playing time is obtained with smaller hardware. This is useful in ENG (electronic news gathering) and consumer devices.
c Tolerances can be relaxed. With fewer data to record, storage density can be reduced making equipment which is more resistant to adverse environments and which requires less maintenance.
d In transmission systems, compression allows a reduction in bandwidth which will generally result in a reduction in cost. This may make possible a service which would be impracticable without it.
e If a given bandwidth is available to an uncompressed signal, compression allows faster than real-time transmission in the same bandwidth.
f If a given bandwidth is available, compression allows a better-quality signal in the same bandwidth.

1.3 MPEG-1, 2, 4 and H.264 contrasted

The first compression standard for audio and video was MPEG-1.[1,2] Although many applications have been found, MPEG-1 was basically designed to allow moving pictures and sound to be encoded into the bit rate of an audio Compact Disc. The resultant Video-CD was quite successful but has now been superseded by DVD. In order to meet the low bit requirement, MPEG-1 downsampled the images heavily as well as using picture rates of only 24–30 Hz and the resulting quality was moderate.

The subsequent MPEG-2 standard was considerably broader in scope and of wider appeal.[3] For example, MPEG-2 supports interlace and HD whereas MPEG-1 did not. MPEG-2 has become very important because it has been chosen as the compression scheme for both DVB (digital video broadcasting) and DVD (digital video disk/digital versatile disk). Developments in standardizing scaleable and multi-resolution compression which would have become MPEG-3 were ready by the time MPEG-2 was ready to be standardized and so this work was incorporated into MPEG-2, and as a result there is no MPEG-3 standard.

MPEG-4[4] uses further coding tools with additional complexity to achieve higher compression factors than MPEG-2. In addition to more efficient coding of video, MPEG-4 moves closer to computer graphics applications. In the more complex Profiles, the MPEG-4 decoder effectively becomes a rendering processor and the compressed bitstream describes three-dimensional shapes and surface texture. It is to be expected that MPEG-4 will become as important to Internet and wireless delivery as MPEG-2 has become in DVD and DVB.

The MPEG-4 standard is extremely wide ranging and it is unlikely that a single decoder will ever be made that can handle every possibility. Many of the graphics applications of MPEG-4 are outside telecommunications requirements. In 2001 the ITU (International Telecommunications Union) Video Coding Experts Group (VCEG) joined with ISO MPEG to form the Joint Video Team (JVT). The resulting standard is variously known as AVC (advanced video coding), H.264 or MPEG-4 Part 10. This standard further refines the video coding aspects of MPEG-4, which were themselves refinements of MPEG-2, to produce a coding scheme having the same applications as MPEG-2 but with higher performance.

To avoid tedium, in cases where the term MPEG is used in this book without qualification, it can be taken to mean MPEG-1, 2, 4 or H.264. Where a specific standard is being contrasted it will be made clear.

1.4 Some applications of compression

The applications of audio and video compression are limitless and the ISO has done well to provide standards which are appropriate to the wide range of possible compression products.

MPEG coding embraces video pictures from the tiny screen of a videophone to the high-definition images needed for electronic cinema. Audio coding stretches from speech-grade mono to multichannel surround sound.

Figure 1.3 shows the use of a codec with a recorder. The playing time of the medium is extended in proportion to the compression factor. In the

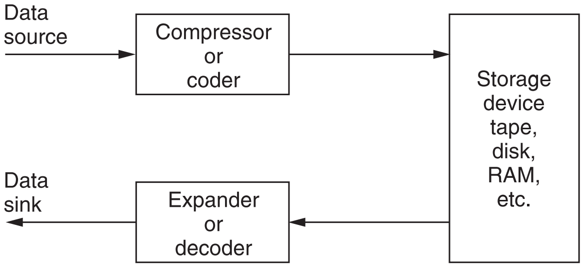


Figure 1.3 Compression can be used around a recording medium. The storage capacity may be increased or the access time reduced according to the application.

case of tapes, the access time is improved because the length of tape needed for a given recording is reduced and so it can be rewound more quickly. In the case of DVD (digital video disk aka digital versatile disk) the challenge was to store an entire movie on one 12 cm disk. The storage density available with today's optical disk technology is such that consumer recording of conventional uncompressed video would be out of the question.

In communications, the cost of data links is often roughly proportional to the data rate and so there is simple economic pressure to use a high compression factor. However, it should be borne in mind that implementing the codec also has a cost which rises with compression factor and so a degree of compromise will be inevitable.

In the case of video-on-demand, technology exists to convey full bandwidth video to the home, but to do so for a single individual at the moment would be prohibitively expensive. Without compression, HDTV (high-definition television) requires too much bandwidth. With compression, HDTV can be transmitted to the home in a similar bandwidth to an existing analog SDTV channel. Compression does not make video-on-demand or HDTV possible; it makes them economically viable.

In workstations designed for the editing of audio and/or video, the source material is stored on hard disks for rapid access. Whilst top-grade systems may function without compression, many systems use compression to offset the high cost of disk storage. In some systems a compressed version of the top-grade material may also be stored for browsing purposes.

When a workstation is used for *off-line* editing, a high compression factor can be used and artifacts will be visible in the picture. This is of no consequence as the picture is only seen by the editor who uses it to make an EDL (edit decision list) which is no more than a list of actions and the timecodes at which they occur. The original uncompressed material is then *conformed* to the EDL to obtain a high-quality edited work. When *on-line* editing is being performed, the output of the workstation is the finished product and clearly a lower compression factor will have to be used. Perhaps it is in broadcasting where the use of compression will have its greatest impact. There is only one electromagnetic spectrum and pressure from other services such as cellular telephones makes efficient use of bandwidth mandatory. Analog television broadcasting is an old technology and makes very inefficient use of bandwidth. Its replacement by a compressed digital transmission is inevitable for the practical reason that the bandwidth is needed elsewhere.

Fortunately in broadcasting there is a mass market for decoders and these can be implemented as low-cost integrated circuits. Fewer encoders are needed and so it is less important if these are expensive. Whilst the cost of digital storage goes down year on year, the cost of the electromagnetic spectrum goes up. Consequently in the future the pressure to use compression in recording will ease whereas the pressure to use it in radio communications will increase.

1.5 Lossless and perceptive coding

Although there are many different coding techniques, all of them fall into one or other of these categories. In *lossless* coding, the data from the expander are identical bit-for-bit with the original source data. The so-called 'stacker' programs which increase the apparent capacity of disk drives in personal computers use lossless codecs. Clearly with computer programs the corruption of a single bit can be catastrophic. Lossless coding is generally restricted to compression factors of around 2:1.

It is important to appreciate that a lossless coder cannot guarantee a particular compression factor and the communications link or recorder used with it must be able to function with the variable output data rate. Source data which result in poor compression factors on a given codec are described as *difficult*. It should be pointed out that the difficulty is often a function of the codec. In other words data which one codec finds difficult may not be found difficult by another. Lossless codecs can be included in bit-error-rate testing schemes. It is also possible to cascade or *concatenate* lossless codecs without any special precautions.

Higher compression factors are only possible with *lossy* coding in which data from the expander are not identical bit-for-bit with the source data and as a result comparing the input with the output is bound to reveal differences. Lossy codecs are not suitable for computer data, but are used in MPEG as they allow greater compression factors than lossless codecs. Successful lossy codecs are those in which the errors are arranged so that a human viewer or listener finds them subjectively difficult to detect. Thus lossy codecs must be based on an understanding of psycho-acoustic and psycho-visual perception and are often called *perceptive* codes.

In perceptive coding, the greater the compression factor required, the more accurately must the human senses be modelled. Perceptive coders can be forced to operate at a fixed compression factor. This is convenient for practical transmission applications where a fixed data rate is easier to handle than a variable rate. The result of a fixed compression factor is that the subjective quality can vary with the 'difficulty' of the input material. Perceptive codecs should not be concatenated indiscriminately especially if they use different algorithms. As the reconstructed signal from a perceptive codec is not bit-for-bit accurate, clearly such a codec cannot be included in any bit error rate testing system as the coding differences would be indistinguishable from real errors.

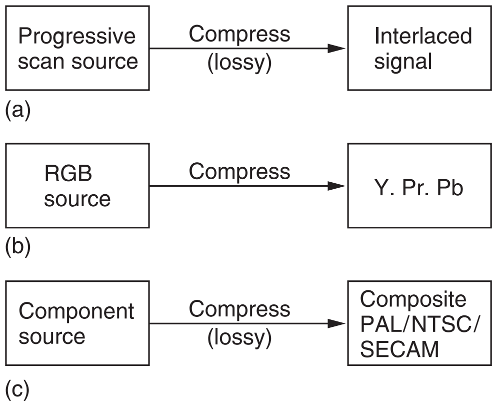


Figure 1.4 Compression is as old as television. (a) Interlace is a primitive way of halving the bandwidth. (b) Colour difference working invisibly reduces colour resolution. (c) Composite video transmits colour in the same bandwidth as monochrome.

Although the adoption of digital techniques is recent, compression itself is as old as television. Figure 1.4 shows some of the compression techniques used in traditional television systems.

Most video signals employ a non-linear relationship between brightness and the signal voltage which is known as gamma. Gamma is a perceptive coding technique which depends on the human sensitivity to video noise being a function of the brightness. The use of gamma allows the same subjective noise level with an eight-bit system as would be achieved with a fourteen-bit linear system.

One of the oldest techniques is interlace, which has been used in analog television from the very beginning as a primitive way of reducing bandwidth. As will be seen in Chapter 5, interlace is not without its problems, particularly in motion rendering. MPEG-2 supports interlace simply because legacy interlaced signals exist and there is a requirement to compress them. This should not be taken to mean that it is a good idea.

The generation of colour difference signals from *RGB* in video represents an application of perceptive coding. The human visual system (HVS) sees no change in quality although the bandwidth of the colour difference signals is reduced. This is because human perception of detail in colour changes is much less than in brightness changes. This approach is sensibly retained in MPEG.

Composite video systems such as PAL, NTSC and SECAM are all analog compression schemes which embed a subcarrier in the luminance signal so that colour pictures are available in the same bandwidth as monochrome. In comparison with a linear-light progressive scan *RGB* picture, gamma-coded interlaced composite video has a compression factor of about 10:1.

In a sense MPEG-2 can be considered to be a modern digital equivalent of analog composite video as it has most of the same attributes. For example, the eight-field sequence of the PAL subcarrier which makes editing difficult has its equivalent in the GOP (group of pictures) of MPEG.

1.6 Compression principles

In a PCM digital system the bit rate is the product of the sampling rate and the number of bits in each sample and this is generally constant. Nevertheless the *information* rate of a real signal varies.

One definition of information is that it is the unpredictable or surprising element of data. Newspapers are a good example of information because they only mention items which are surprising. Newspapers never carry items about individuals who have *not* been involved in an accident as this is the normal case. Consequently the phrase 'no news is good news' is remarkably true because if an information channel exists but nothing has been sent then it is most likely that nothing remarkable has happened.

The unpredictability of the punch line is a useful measure of how funny a joke is. Often the build-up paints a certain picture in the listener's imagination, which the punch line destroys utterly. One of the author's favourites is the one about the newly married couple who didn't know the difference between putty and petroleum jelly – their windows fell out. The difference between the information rate and the overall bit rate is known as the redundancy. Compression systems are designed to eliminate as much of that redundancy as practicable or perhaps affordable. One way in which this can be done is to exploit statistical predictability in signals. The information content or *entropy* of a sample is a function of how different it is from the predicted value. Most signals have some degree of predictability. A sine wave is highly predictable, because all cycles look the same. According to Shannon's theory, any signal which is totally predictable carries no information. In the case of the sine wave this is clear because it represents a single frequency and so has no bandwidth.

In all real signals, at least part of the signal is obvious from what has gone before or, in some cases, what may come later. A suitable decoder can predict some of the obvious part of the signal so that only the remainder has to be sent. Figure 1.5 shows a codec based on prediction. The decoder contains a predictor that attempts to anticipate the next string of data from what has gone before. If the characteristics of the decoder's predictor are known, the transmitter can omit parts of the message that the receiver has the ability to re-create. It should be clear

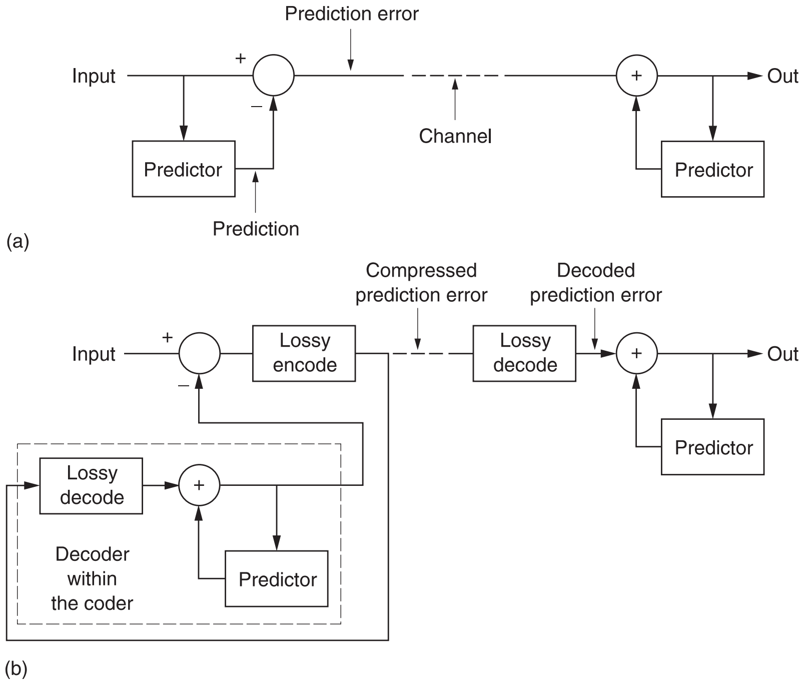


Figure 1.5 (a) A predictive codec has identical predictors in both encoder and decoder. The prediction error, or residual, is sent to the decoder which uses it perfectly to cancel the prediction error. Such a codec is lossless. (b) If the residual is subject to further lossy compression, this must take place within the encoder's prediction loop so that both predictors can track.

that all encoders using prediction must contain the same predictor as the decoder, or at least a model of it.

In Figure 1.5(a) it will be seen that the predictors in the encoder and the decoder operate with identical inputs: namely the output of the encoder. If the predictors are identical, the predictions will be identical. Thus the encoder can subtract its own prediction from the actual input to obtain a prediction error, or *residual*, in the knowledge that the decoder can add that residual thereby cancelling the prediction error.

This is a powerful technique because if all of the residual is transmitted, the codec is completely lossless. In practical codecs, lossless prediction may be combined with lossy tools. This has to be done with care. Figure 1.5(b) shows a predictive coder in which the residual is subject to further lossy coding. Note that the lossy coder is inside the encoder's prediction loop so that once more the predictors in the encoder and the decoder see the same signal. If this is not done the outputs of the two predictors will drift apart.

Clearly a signal such as noise is completely unpredictable and as a result all codecs find noise *difficult*. The most efficient way of coding noise is PCM. A codec which is designed using the statistics of real material should not be tested with random noise because it is not a representative test. Second, a codec which performs well with clean source material may perform badly with source material containing superimposed noise. Most practical compression units require some form of pre-processing before the compression stage proper and appropriate noise reduction should be incorporated into the pre-processing if noisy signals are anticipated. It will also be necessary to restrict the degree of compression applied to noisy signals.

All real signals fall part-way between the extremes of total predictability and total unpredictability or noisiness. If the bandwidth (set by the sampling rate) and the dynamic range (set by the wordlength) of the transmission system are used to delineate an area, this sets a limit on the information capacity of the system. Figure 1.6(a) shows that most real

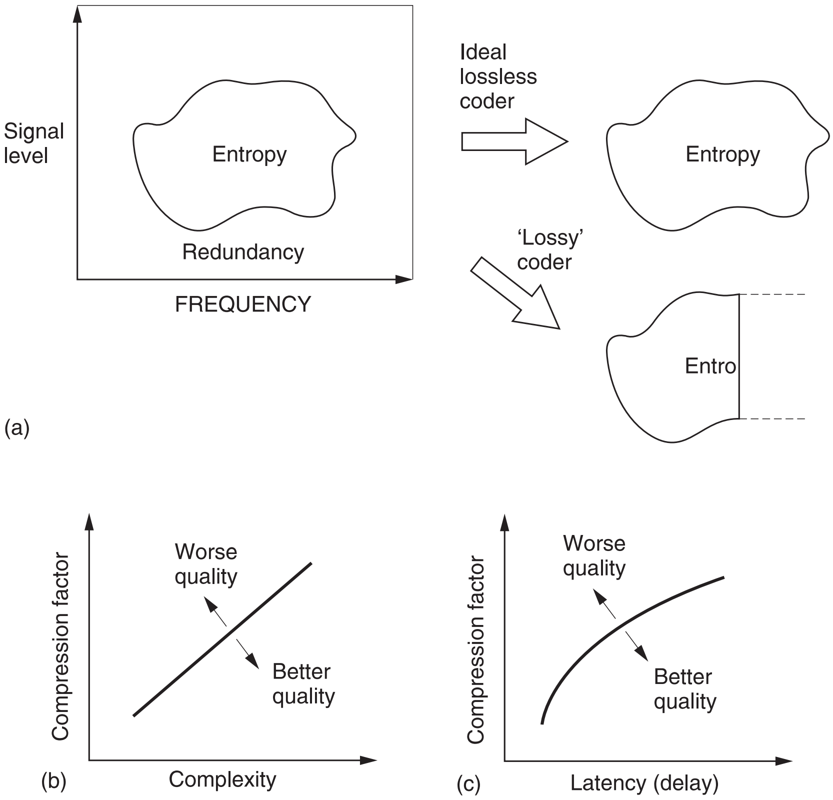


Figure 1.6 (a) A perfect coder removes only the redundancy from the input signal and results in subjectively lossless coding. If the remaining entropy is beyond the capacity of the channel some of it must be lost and the codec will then be lossy. An imperfect coder will also be lossy as it fails to keep all entropy. (b) As the compression factor rises, the complexity must also rise to maintain quality. (c) High compression factors also tend to increase latency or delay through the system.

signals only occupy part of that area. The signal may not contain all frequencies, or it may not have full dynamics at certain frequencies.

Entropy can be thought of as a measure of the actual area occupied by the signal. This is the area that *must* be transmitted if there are to be no subjective differences or *artifacts* in the received signal. The remaining area is called the *redundancy* because it adds nothing to the information conveyed. Thus an ideal coder could be imagined which miraculously sorts out the entropy from the redundancy and only sends the former. An ideal decoder would then re-create the original impression of the information quite perfectly. In this ideal case, the entropy and the residual would be the same. As the ideal is approached, the coder complexity and the latency or delay both rise. In real coders, the ideal is not reached and the residual must be larger than the entropy. Figure 1.6(b) shows how complexity increases with compression factor. The additional complexity of H.264 over MPEG-2 is obvious from this. Figure 1.6(c) shows how increasing the codec latency can improve the compression factor.

Nevertheless moderate coding gains that only remove redundancy need not cause artifacts and result in systems which are described as *subjectively lossless*. If the channel capacity is not sufficient for that, then the coder will have to discard some of the entropy and with it useful information. Larger coding gains which remove some of the entropy must result in artifacts. It will also be seen from Figure 1.6 that an imperfect coder will fail to separate the redundancy and may discard entropy instead, resulting in artifacts at a sub-optimal compression factor.

It should be clear from the above that if quality is an issue, the highest performance will be obtained using codecs in which the prediction is as powerful as possible. When this is done the residual will be smaller and so the degree of lossy coding needed to achieve a target bit rate will be reduced.

Obviously it is necessary to provide a channel that could accept whatever entropy the coder extracts in order to have transparent quality. A single variable rate transmission or recording channel is traditionally unpopular with channel providers, although newer systems such as ATM support variable rate. Digital transmitters used in DVB or ATSC have a fixed bit rate. The variable rate requirement can be overcome by combining several compressed channels into one constant rate transmission in a way which flexibly allocates data rate between the channels. Provided the material is unrelated, the probability of all channels reaching peak entropy at once is very small and so those channels which are at one instant passing easy material will make available transmission capacity for those channels which are handling difficult material. This is the principle of statistical multiplexing.

Where the same type of source material is used consistently, e.g. English text, then it is possible to perform a statistical analysis on the frequency with which particular letters are used. Variable-length coding is used in which frequently used letters are allocated short codes and letters which occur infrequently are allocated long codes. This results in a lossless code. The well-known Morse code used for telegraphy is an example of this approach. The letter e is the most frequent in English and is sent with a single dot. An infrequent letter such as z is allocated a long complex pattern. It should be clear that codes of this kind which rely on a prior knowledge of the statistics of the signal are only effective with signals actually having those statistics. If Morse code is used with another language, the transmission becomes significantly less efficient because the statistics are quite different; the letter z, for example, is quite common in Czech.

The Huffman code[5] is also one which is designed for use with a data source having known statistics. The probability of the different code values to be transmitted is studied, and the most frequent codes are arranged to be transmitted with short wordlength symbols. As the probability of a code value falls, it will be allocated longer wordlength. The Huffman code is used in conjunction with a number of compression techniques and is shown in Figure 1.7.

The input or *source* codes are assembled in order of descending probability. The two lowest probabilities are distinguished by a single

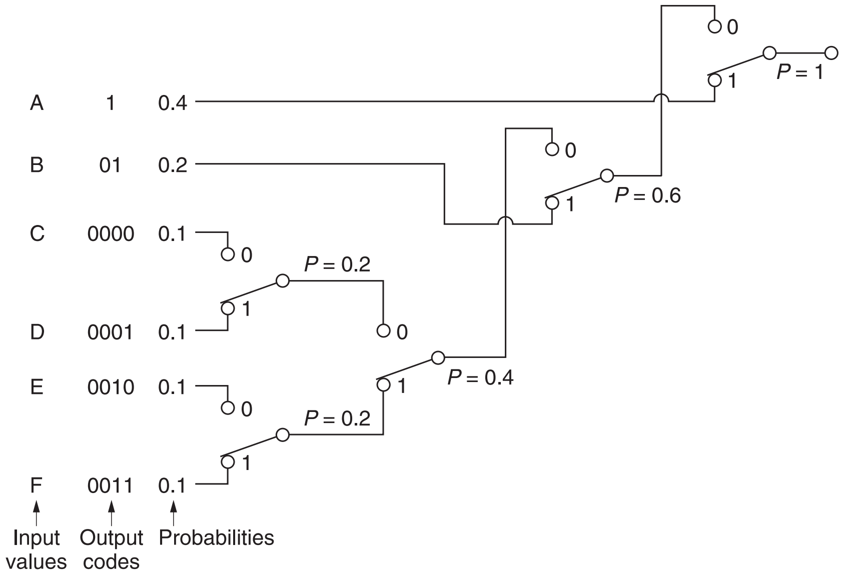


Figure 1.7 The Huffman code achieves compression by allocating short codes to frequent values. To aid deserializing the short codes are not prefixes of longer codes.

code bit and their probabilities are combined. The process of combining probabilities is continued until unity is reached and at each stage a bit is used to distinguish the path. The bit will be a zero for the most probable path and one for the least. The compressed output is obtained by reading the bits which describe which path to take going from right to left.

In the case of computer data, there is no control over the data statistics. Data to be recorded could be instructions, images, tables, text files and so on; each having their own code value distributions. In this case a coder relying on fixed source statistics will be completely inadequate. Instead a system is used which can learn the statistics as it goes along. The Lempel–Ziv–Welch (LZW) lossless codes are in this category. These codes build up a conversion table between frequent long source data strings and short transmitted data codes at both coder and decoder and initially their compression factor is below unity as the contents of the conversion tables are transmitted along with the data. However, once the tables are established, the coding gain more than compensates for the initial loss. In some applications, a continuous analysis of the frequency of code selection is made and if a data string in the table is no longer being used with sufficient frequency it can be deselected and a more common string substituted.

Lossless codes are less common for audio and video coding where perceptive codes are permissible. The perceptive codes often obtain a coding gain by shortening the wordlength of the data representing the signal waveform. This must increase the noise level and the trick is to ensure that the resultant noise is placed at frequencies where human senses are least able to perceive it. As a result although the received signal is measurably different from the source data, it can *appear* the same to the human listener or viewer at moderate compression factors. As these codes rely on the characteristics of human sight and hearing, they can only be fully tested subjectively.

The compression factor of such codes can be set at will by choosing the wordlength of the compressed data. Whilst mild compression will be undetectable, with greater compression factors, artifacts become noticeable. Figure 1.6 shows that this is inevitable from entropy considerations.

1.7 Video compression

Video signals exist in four dimensions: these are the attributes of the pixel, the horizontal and vertical spatial axes and the time axis. Compression can be applied in any or all of those four dimensions. MPEG assumes an eight-bit colour difference signal as the input, requiring rounding if the source is ten bit. The sampling rate of the

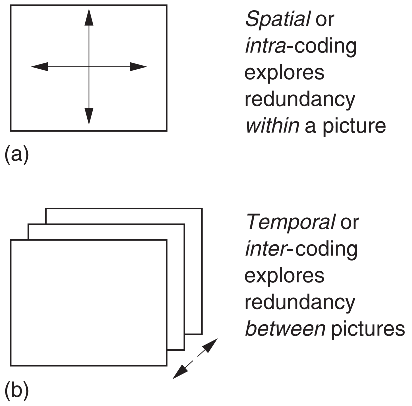


Figure 1.8 (a) Spatial or intra-coding works on individual images. (b) Temporal or inter-coding works on successive images.

colour signals is less than that of the luminance. This is done by downsampling the colour samples horizontally and generally vertically as well. Essentially an MPEG system has three parallel simultaneous channels, one for luminance and two colour difference, which after coding are multiplexed into a single bitstream.

Figure 1.8(a) shows that when individual pictures are compressed without reference to any other pictures, the time axis does not enter the process which is therefore described as *intra-coded* (intra = within) compression. The term *spatial coding* will also be found. It is an advantage of intra-coded video that there is no restriction to the editing which can be carried out on the picture sequence. As a result compressed VTRs such as Digital Betacam, DVC and D-9 use spatial coding. Cut editing may take place on the compressed data directly if necessary. As spatial coding treats each picture independently, it can employ certain techniques developed for the compression of still pictures. The ISO JPEG (Joint Photographic Experts Group) compression standards[6,7] are in this category. Where a succession of JPEG coded images are used for television, the term 'Motion JPEG' will be found.

Greater compression factors can be obtained by taking account of the redundancy from one picture to the next. This involves the time axis, as Figure 1.8(b) shows, and the process is known as *inter-coded* (inter = between) or *temporal* compression.

Temporal coding allows a higher compression factor, but has the disadvantage that an individual picture may exist only in terms of the differences from a previous picture. Clearly editing must be undertaken with caution and arbitrary cuts simply cannot be performed on the MPEG bitstream. If a previous picture is removed by an edit, the difference data will then be insufficient to re-create the current picture.

1.7.1 Intra-coded compression

Intra-coding works in three dimensions on the horizontal and vertical spatial axes and on the sample values. Analysis of typical television pictures reveals that whilst there is a high spatial frequency content due to detailed areas of the picture, there is a relatively small amount of energy at such frequencies. Often pictures contain sizeable areas in which the same or similar pixel values exist. This gives rise to low spatial frequencies. The average brightness of the picture results in a substantial zero frequency component. Simply omitting the high-frequency components is unacceptable as this causes an obvious softening of the picture.

A coding gain can be obtained by taking advantage of the fact that the amplitude of the spatial components falls with frequency. It is also possible to take advantage of the eye's reduced sensitivity to noise in high spatial frequencies. If the spatial frequency spectrum is divided into frequency bands the high-frequency bands can be described by fewer bits not only because their amplitudes are smaller but also because more noise can be tolerated. The wavelet transform (MPEG-4 only) and the discrete cosine transform used in JPEG and MPEG-1, MPEG-2 and MPEG-4 allow two-dimensional pictures to be described in the frequency domain and these are discussed in Chapter 3.

1.7.2 Inter-coded compression

Inter-coding takes further advantage of the similarities between successive pictures in real material. Instead of sending information for each picture separately, inter-coders will send the difference between the previous picture and the current picture in a form of differential coding. Figure 1.9 shows the principle. A picture store is required at the coder to allow comparison to be made between successive pictures and a similar store is required at the decoder to make the previous picture available. The difference data may be treated as a picture itself and subjected to some form of transform-based spatial compression.

The simple system of Figure 1.9(a) is of limited use as in the case of a transmission error, every subsequent picture would be affected. Channel switching in a television set would also be impossible. In practical systems a modification is required. One approach is the so-called 'leaky predictor' in which the next picture is predicted from a limited number of previous pictures rather than from an indefinite number. As a result errors cannot propagate indefinitely. The approach used in MPEG is that periodically some absolute picture data are transmitted in place of difference data.

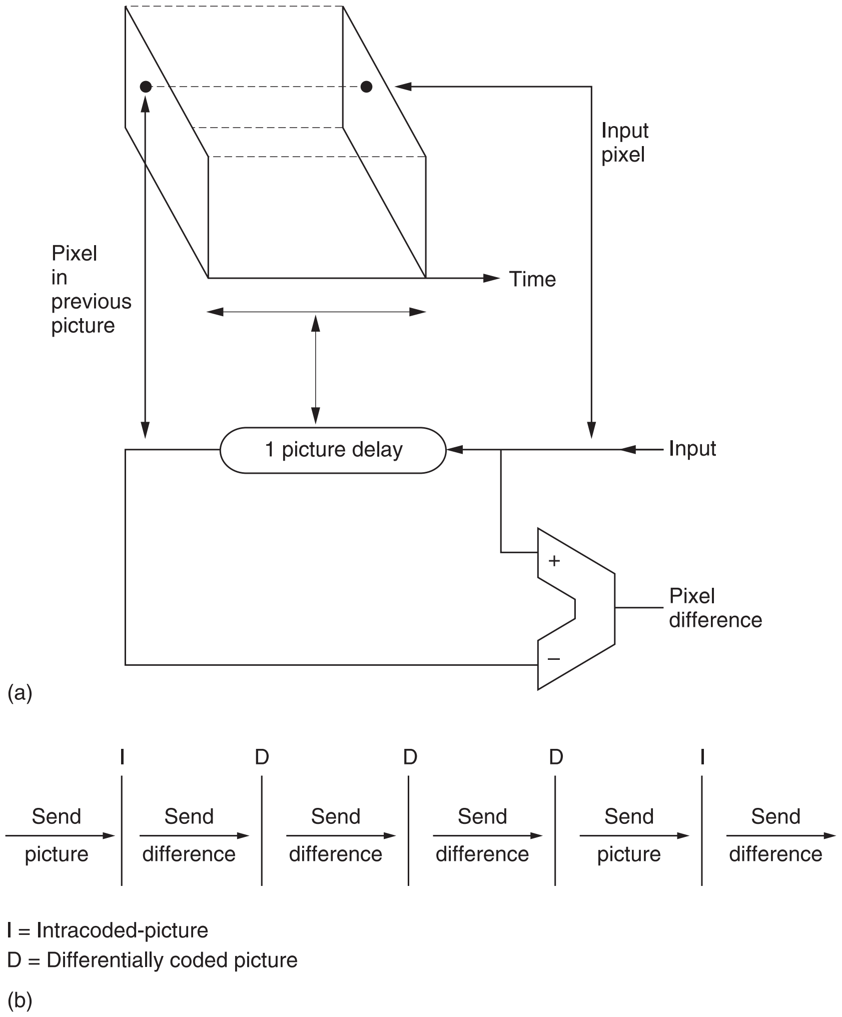


Figure 1.9 An inter-coded system (a) uses a delay to calculate the pixel differences between successive pictures. To prevent error propagation, intra-coded pictures (b) may be used periodically.

Figure 1.9(b) shows that absolute picture data, known as *I* or *intra pictures* are interleaved with pictures which are created using difference data, known as *P* or *predicted* pictures. The *I* pictures require a large amount of data, whereas the *P* pictures require fewer data. As a result the instantaneous data rate varies dramatically and buffering has to be used to allow a constant transmission rate. The leaky predictor needs less buffering as the compression factor does not change so much from picture to picture.

The *I* picture and all of the *P* pictures prior to the next *I* picture are called a group of pictures (GOP). For a high compression factor, a large number of *P* pictures should be present between *I* pictures, making a long GOP. However, a long GOP delays recovery from a transmission error. The compressed bitstream can only be edited at *I* pictures as shown.

In the case of moving objects, although their appearance may not change greatly from picture to picture, the data representing them on a fixed sampling grid will change and so large differences will be generated between successive pictures. It is a great advantage if the effect of motion can be removed from difference data so that they only reflect the changes in appearance of a moving object since a much greater coding gain can then be obtained. This is the objective of motion compensation.

1.7.3 Introduction to motion compensation

In real television program material objects move around before a fixed camera or the camera itself moves. Motion compensation is a process which effectively measures motion of objects from one picture to the next so that it can allow for that motion when looking for redundancy between pictures. Figure 1.10 shows that moving pictures can be expressed in a three-dimensional space which results from the screen area moving along the time axis. In the case of still objects, the only motion is along the time axis. However, when an object moves, it does so along the *optic flow axis* which is not parallel to the time axis. The optic

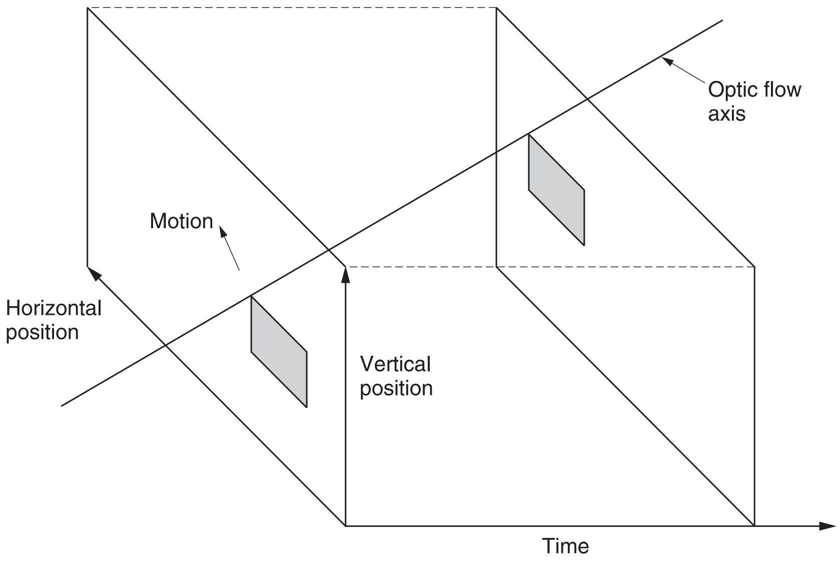


Figure 1.10 Objects travel in a three-dimensional space along the optic flow axis which is only parallel to the time axis if there is no movement.

flow axis is the locus of a point on a moving object as it takes on various screen positions.

It will be clear that the data values representing a moving object change with respect to the time axis. However, looking along the optic flow axis the appearance of an object only changes if it deforms, moves into shadow or rotates. For simple translational motions the data representing an object are highly redundant with respect to the optic flow axis. Thus if the optic flow axis can be located, coding gain can be obtained in the presence of motion.

A motion-compensated coder works as follows. A reference picture is sent, but is also locally stored so that it can be compared with another picture to find motion vectors for various areas of the picture. The reference picture is then shifted according to these vectors to cancel interpicture motion. The resultant *predicted* picture is compared with the actual picture to produce a *prediction error* also called a *residual*. The prediction error is transmitted with the motion vectors. At the receiver the reference picture is also held in a memory. It is shifted according to the transmitted motion vectors to re-create the predicted picture and then the prediction error is added to it to re-create the original.

In prior compression schemes the predicted picture followed the reference picture. In MPEG this is not the case. Information may be brought back from a later picture or forward from an earlier picture as appropriate.

1.7.4 Film-originated video compression

Film can be used as the source of video signals if a telecine machine is used. The most common frame rate for film is 24 Hz, whereas the field rates of television are 50 Hz and 60 Hz. This incompatibility is patched over in two different ways. In 50 Hz telecine, the film is simply played slightly too fast so that the frame rate becomes 25 Hz. Then each frame is converted into two television fields giving the correct 50 Hz field rate. In 60 Hz telecine, the film travels at the correct speed, but alternate frames are used to produce two fields then three fields. The technique is known as 3:2 pulldown. In this way two frames produce five fields and so the correct 60 Hz field rate results. The motion portrayal of telecine is not very good as moving objects judder, especially in 60 Hz systems. Figure 1.11 shows how the optic flow is portrayed in film-originated video.

When film-originated video is input to a compression system, the disturbed optic flow will play havoc with the motion-compensation system. In a 50 Hz system there appears to be no motion between the two fields which have originated from the same film frame, whereas between

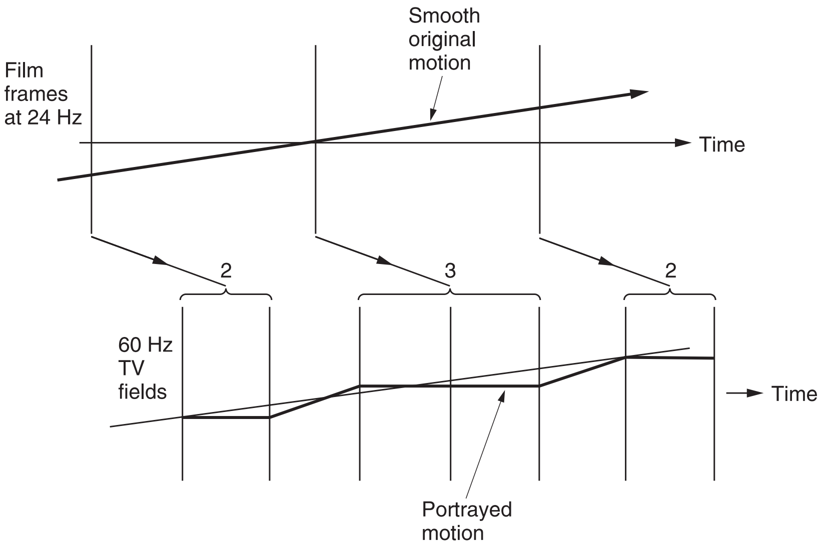


Figure 1.11 Telecine machines must use 3:2 pulldown to produce 60 Hz field rate video.

the next two fields large motions will exist. In 60 Hz systems, the motion will be zero for three fields out of five.

With such inputs, it is more efficient to adopt a different processing mode which is based upon the characteristics of the original film. Instead of attempting to manipulate fields of video, the system de-interlaces pairs of fields in order to reconstruct the original film frames. This can be done by a fairly simple motion detector. When substantial motion is measured between successive fields in the output of a telecine, this is taken to mean that the fields have come from different film frames. When negligible motion is detected between fields, this is taken to indicate that the fields have come from the same film frame.

In 50 Hz video it is quite simple to find the sequence and produce de-interlaced frames at 25 Hz. In 60 Hz 3:2 pulldown video the problem is slightly more complex because it is necessary to locate the frames in which three fields are output so that the third field can be discarded, leaving, once more, de-interlaced frames at 25 Hz. Whilst it is relatively straightforward to lock on to the 3:2 sequence with direct telecine output signals, if the telecine material has been edited on videotape the 3:2 sequence may contain discontinuities. In this case it is necessary to provide a number of field stores in the de-interlace unit so that a series of fields can be examined to locate the edits. Once telecine video has been de-interlaced back to frames, intra- and inter-coded compression can be employed using frame-based motion compensation.

MPEG transmissions include flags that tell the decoder the origin of the material. Material originating at 24 Hz but converted to interlaced video does not have the motion attributes of interlace because the lines in two fields have come from the same point on the time axis. Two fields can be combined to create a progressively scanned frame. In the case of 3:2 pulldown material, the third field need not be sent at all as the decoder can easily repeat a field from memory. As a result the same compressed film material can be output at 50 or 60 Hz as required.

Recently conventional telecine machines have been superseded by the *datacine* which scans each film frame into a pixel array which can be made directly available to the MPEG encoder without passing through an intermediate digital video standard. Datacines are used extensively for mastering DVDs from film stock.

1.8 Introduction to MPEG-1

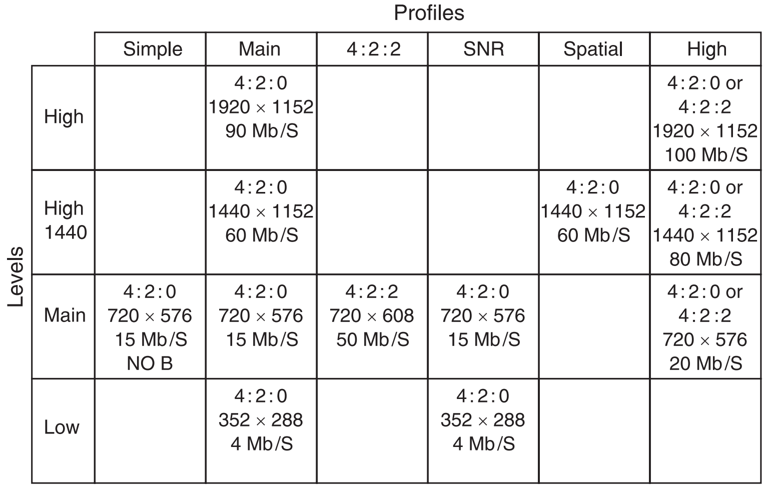
As mentioned above, the intention of MPEG-1 is to deliver video and audio at the same bit rate as a conventional audio CD. As the bit rate was a given, this was achieved by subsampling to half the definition of conventional television. In order to have a constant input bit rate irrespective of the frame rate, 25 Hz systems have a picture size of 352 × 288 pixels whereas 30 Hz systems have a picture size of 352 × 240 pixels. This is known as *common intermediate format* (CIF). If the input is conventional interlaced video, CIF can be obtained by discarding alternate fields and downsampling the remaining active lines by a factor of two.

As interlaced systems have very poor vertical resolution, downsampling to CIF actually does little damage to still images, although the very low picture rates damage motion portrayal.

Although MPEG-1 appeared rather rough on screen, this was due to the very low bit rate. It is more important to appreciate that MPEG-1 introduced the great majority of the coding tools which would continue to be used in MPEG-2 and MPEG-4. These included an elementary stream syntax, bidirectional motion-compensated coding, buffering and rate control. Many of the spatial coding principles of MPEG-1 were taken from JPEG. MPEG-1 also specified audio compression of up to two channels.

1.9 MPEG-2: Profiles and Levels

MPEG-2 builds upon MPEG-1 by adding interlace capability as well as a greatly expanded range of picture sizes and bit rates. The use of scaleable systems is also addressed, along with definitions of how multiple MPEG bitstreams can be multiplexed. As MPEG-2 is an extension of MPEG-1, it is easy for MPEG-2 decoders to handle MPEG-1 data. In a sense an



		Profiles					
		Simple	Main	4:2:2	SNR	Spatial	High
Levels	High		4:2:0 1920 × 1152 90 Mb/S				4:2:0 or 4:2:2 1920 × 1152 100 Mb/S
	High 1440		4:2:0 1440 × 1152 60 Mb/S			4:2:0 1440 × 1152 60 Mb/S	4:2:0 or 4:2:2 1440 × 1152 80 Mb/S
	Main	4:2:0 720 × 576 15 Mb/S NO B	4:2:0 720 × 576 15 Mb/S	4:2:2 720 × 608 50 Mb/S	4:2:0 720 × 576 15 Mb/S		4:2:0 or 4:2:2 720 × 576 20 Mb/S
	Low		4:2:0 352 × 288 4 Mb/S		4:2:0 352 × 288 4 Mb/S		

Figure 1.12 Profiles and Levels in MPEG-2. See text for details.

MPEG-1 bitstream is an MPEG-2 bitstream which has a restricted vocabulary and so can be readily understood by an MPEG-2 decoder.

MPEG-2 has too many applications to solve with a single standard and so it is subdivided into Profiles and Levels. Put simply a Profile describes a degree of complexity whereas a Level describes the picture size or resolution which goes with that Profile. Not all Levels are supported at all Profiles. Figure 1.12 shows the available combinations. In principle there are twenty-four of these, but not all have been defined. An MPEG-2 decoder having a given Profile and Level must also be able to decode lower Profiles and Levels.

The simple Profile does not support bidirectional coding and so only *I* and *P* pictures will be output. This reduces the coding and decoding delay and allows simpler hardware. The simple Profile has only been defined at Main Level (SP ML).

The Main Profile is designed for a large proportion of uses. The Low Level uses a low resolution input having only 352 pixels per line. The majority of broadcast applications will require the MP ML (Main Profile at Main Level) subset of MPEG which supports SDTV (standard definition television). The High-1440 Level is a high-definition scheme which doubles the definition compared to Main Level. The High Level not only doubles the resolution but maintains that resolution with 16:9 format by increasing the number of horizontal samples from 1440 to 1920.

In compression systems using spatial transforms and requantizing it is possible to produce scaleable signals. A scaleable process is one in which the input results in a main signal and a 'helper' signal. The main signal can be decoded alone to give a picture of a certain quality, but if the

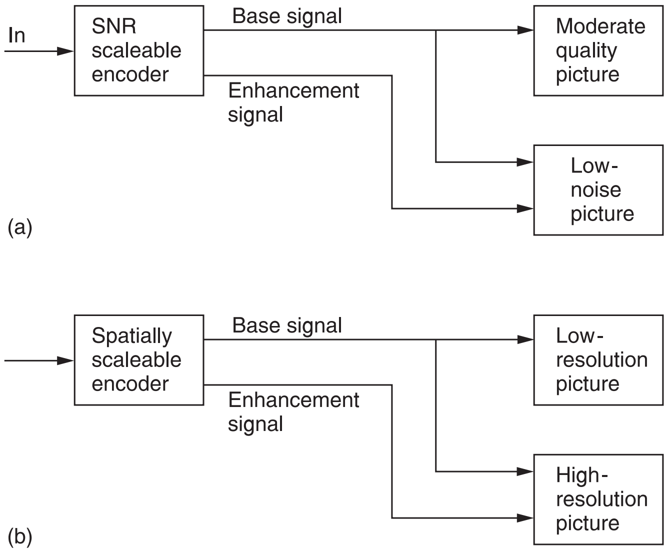


Figure 1.13 (a) An SNR scaleable encoder produces a 'noisy' signal and a noise cancelling signal. (b) A spatially scaleable encoder produces a low-resolution picture and a resolution-enhancing picture.

information from the helper signal is added some aspect of the quality can be improved.

Figure 1.13(a) shows that in a conventional MPEG coder, by heavily requantizing coefficients a picture with moderate signal-to-noise ratio results. If, however, that picture is locally decoded and subtracted pixel by pixel from the original, a 'quantizing noise' picture would result. This can be compressed and transmitted as the helper signal. A simple decoder only decodes the main 'noisy' bitstream, but a more complex decoder can decode both bitstreams and combine them to produce a low-noise picture. This is the principle of SNR scaleability.

As an alternative, Figure 1.13(b) shows that by coding only the lower spatial frequencies in an HDTV picture a base bitstream can be made which an SDTV receiver can decode. If the lower definition picture is locally decoded and subtracted from the original picture, a 'definition-enhancing' picture would result. This can be coded into a helper signal.

A suitable decoder could combine the main and helper signals to recreate the HDTV picture. This is the principle of spatial scaleability. The High Profile supports both SNR and spatial scaleability as well as allowing the option of 4:2:2 sampling (see section 2.11).

The 4:2:2 Profile has been developed for improved compatibility with existing digital television production equipment. This allows 4:2:2 working without requiring the additional complexity of using the High Profile. For example, an HP ML decoder must support SNR scaleability which is not a requirement for production.

MPEG-2 increased the number of audio channels possible to five whilst remaining compatible with MPEG-1 audio. MPEG-2 subsequently introduced a more efficient audio coding scheme known as MPEG-2 AAC (advanced audio coding) which is not backwards compatible with the earlier audio coding schemes.

1.10 Introduction to MPEG-4

MPEG-4 introduces a number of new coding tools as shown in Figure 1.14. In MPEG-1 and MPEG-2 the motion compensation is based on regular fixed-size areas of image known as *macroblocks*. Whilst this works well at the designed bit rates, there will always be some inefficiency due to real moving objects failing to align with macroblock boundaries. This will increase the residual bit rate. In MPEG-4, moving objects can be coded as arbitrary shapes. Figure 1.15 shows that a background can be coded quite independently from objects in front of it. Object motion can then be described with vectors and much-reduced residual data.

According to the Profile, objects may be two dimensional, three dimensional and opaque or translucent. The decoder must contain effectively a layering vision mixer which is capable of prioritizing image data as a function of how close it is to the viewer. The picture coding of MPEG-4 is known as texture coding and is more advanced than the MPEG-2 equivalent, using more lossless predictive coding for pixel values, coefficients and vectors.

In addition to motion compensation, MPEG-4 can describe how an object changes its perspective as it moves using a technique called mesh

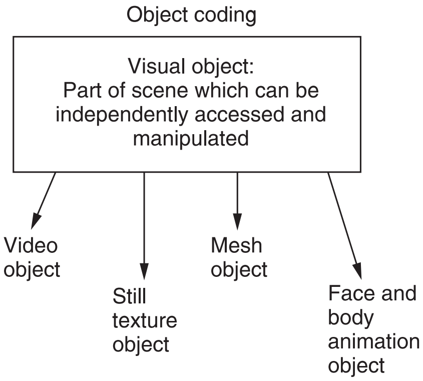

Object coding
Visual object:
Part of scene which can be independently accessed and manipulated
Video object
Still texture object
Mesh object
Face and body animation object

Figure 1.14 MPEG-4 introduces a number of new coding tools over those of earlier MPEG standards. These include object coding, mesh coding, still picture coding and face and body animation.

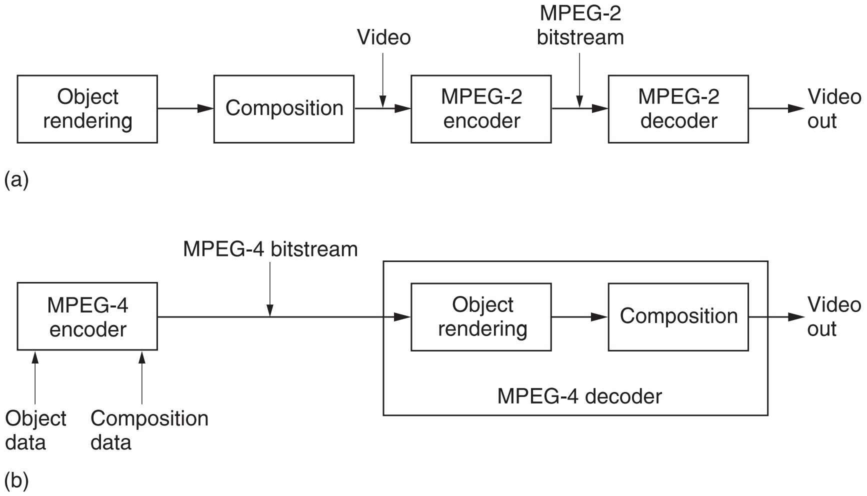


Figure 1.15 (a) In MPEG-1 and MPEG-2, computer graphic images must be rendered to video before coding. (b) In contrast, MPEG-4 may move the rendering process to the decoder, reducing the bit rate needed with the penalty of increased decoder complexity.

coding. By warping another image, the prediction of the present image is improved. MPEG-4 also introduces coding for still images using DCT or wavelets.

Although MPEG-2 supported some scaleability, MPEG-4 also takes this further. In addition to spatial and noise scaleability, MPEG-4 also allows temporal scaleability where a base level bitstream having a certain frame rate may be augmented by an additional enhancement bitstream to produce a decoder output at a higher frame rate. This is important as it allows a way forward from the marginal frame rates of today’s film and television formats whilst remaining backwards compatible with traditional equipment. The comprehensive scaleability of MPEG-4 is equally important in networks where it allows the user the best picture possible for the available bit rate.

MPEG-4 also introduces standards for face and body animation. Specialized vectors allow a still picture of a face and optionally a body to be animated to allow expressions and gestures to accompany speech at very low bit rates. In some senses MPEG-4 has gone upstream of the video signal which forms the input to MPEG-1 and MPEG-2 coders to analyse ways in which the video signal was rendered. Figure 1.15(a) shows that in a system using MPEG-1 and MPEG-2, all rendering and production steps take place before the encoder. Figure 1.15(b) shows that in MPEG-4, some of these steps can take place in the decoder. The advantage is that fewer data need to be transmitted. Some of these data will be rendering instructions which can be very efficient and result in a high compression factor. As a significant part of the rendering takes place

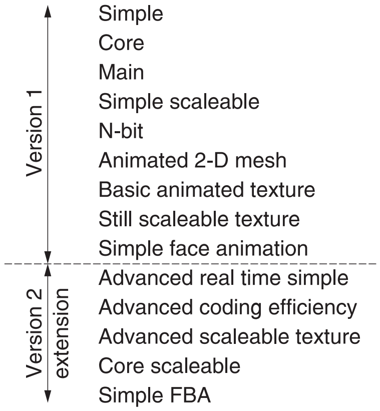


Figure 1.16 The visual object types supported by MPEG-4 in versions 1 and 2.

in the decoder, computer graphics generators can be designed directly to output an MPEG-4 bitstream. In interactive systems such as simulators and video games, inputs from the user can move objects around the screen. The disadvantage is increased decoder complexity, but as the economics of digital processing continues to advance this is hardly a serious concern.

As might be expected, the huge range of coding tools in MPEG-4 is excessive for many applications. As with MPEG-2 this has been dealt with using Profiles and Levels. Figure 1.16 shows the range of Visual Object types in version 1 of MPEG-4 and as expanded in version 2. For each visual object type the coding tools needed are shown. Figure 1.17 shows the relationship between the Visual Profiles and the Visual Object types supported by each Profile. The crossover between computer-generated and natural images is evident in the Profile structure where Profiles 1–5 cover natural images, Profiles 8 and 9 cover rendered images and Profiles 6 and 7 cover hybrid natural/rendered images. It is only possible to give an introduction here and more detail is provided in Chapter 5. MPEG-4 also extends the boundaries of audio coding. The MPEG-2 AAC technique is extended in MPEG-4 by some additional tools. New tools are added which allow operation at very low bit rates for speech applications. Also introduced is the concept of *structured audio* in which the audio waveform is synthesized at the decoder from a bitstream which is essentially a digital musical score.

1.11 Introduction to H.264 (AVC)

The wide range of applications of MPEG-4 are not needed for broadcasting purposes. However, areas of MPEG-4 such as texture

Version 1 visual profiles

	Object types Profiles	Simple	Core	Main	Simple scaleable	N-bit	Animated 2D mesh	Basic animated texture	Scaleable texture	Simple face
1.	Simple	×								
2.	Simple scaleable	×			×					
3.	Core	×	×							
4.	Main	×	×	×					×	
5.	N-bit	×	×			×				
6.	Hybrid	×	×				×	×	×	×
7.	Basic animated texture							×	×	×
8.	Scaleable texture								×	
9.	Simple FA									×

Version 2 visual profiles

	Object types Profiles	Simple	Simple scaleable	Core	Core scaleable	Advanced real time simple	Advanced coding efficiency	Advanced scaleable texture	Simple FBA
V2-1.	Advanced real time simple	×				×			
V2-2.	Core scaleable	×	×	×	×				
V2-3.	Advanced coding efficiency	×		×			×		
V2-4.	Advanced core	×		×				×	
V2-5.	Advanced scaleable texture							×	
V2-6.	Simple FBA								×

Figure 1.17 The visual object types supported by each visual profile of MPEG-4.

coding are directly applicable to, and indeed can be interpreted as extensions of, MPEG-2. Thus a good way of appreciating H.264[8] is to consider that in developing it, every video coding tool of MPEG-4 was considered and if a refinement was possible, albeit with the penalty of increased complexity, then it would be incorporated. Thus, H.264 AVC does not do more than MPEG-2, but it does it better in that a greater coding gain is achieved for the same perceived quality. It should be noted that a significant increase in processing power is needed with AVC in comparison with MPEG-2. As before, H.264 has Profiles and Levels for the same reasons.

1.12 Audio compression

Perceptive coding in audio relies on the principle of auditory masking, which is treated in detail in section 4.1. Masking causes the ear/brain combination to be less sensitive to sound at one frequency in the presence of another at a nearby frequency. If a first tone is present in the input, then it will mask signals of lower level at nearby frequencies. The quantizing of the first tone and of further tones at those frequencies can be made coarser. Fewer bits are needed and a coding gain results. The increased quantizing error is allowable if it is masked by the presence of the first tone.

1.12.1 Sub-band coding

Sub-band coding mimics the frequency analysis mechanism of the ear and splits the audio spectrum into a large number of different bands. Signals in these bands can then be quantized independently. The quantizing error which results is confined to the frequency limits of the band and so it can be arranged to be masked by the program material. The techniques used in Layers I and II of MPEG audio are based on sub-band coding as are those used in DCC (Digital Compact Cassette).

1.12.2 Transform coding

In transform coding the time-domain audio waveform is converted into a frequency domain representation such as a Fourier, discrete cosine or wavelet transform (see Chapter 3). Transform coding takes advantage of the fact that the amplitude or envelope of an audio signal changes relatively slowly and so the coefficients of the transform can be transmitted relatively infrequently. Clearly such an approach breaks down in the presence of transients and adaptive systems are required in practice. Transients cause the coefficients to be updated frequently whereas in stationary parts of the signal such as sustained notes the update rate can be reduced. Discrete cosine transform (DCT) coding is used in Layer III of MPEG audio and in the compression system of the Sony MiniDisc.

1.12.3 Predictive coding

As seen above, in a predictive coder there are two identical predictors, one in the coder and one in the decoder. Their job is to examine a run of previous data values and to extrapolate forward to estimate or predict what the next value will be. This is subtracted from the *actual* next code value at the encoder to produce a prediction error which is transmitted. The decoder then adds the prediction error to its own prediction to obtain the output code value again.

Prediction can be used in the time domain, where sample values are predicted, or in the frequency domain where coefficient values are predicted. Time-domain predictive coders work with a short encode and decode delay and are useful in telephony where a long loop delay causes problems. Frequency prediction is used in AC-3 and MPEG AAC.

1.13 MPEG bitstreams

MPEG supports a variety of bitstream types for various purposes and these are shown in Figure 1.18. The output of a single compressor (video or audio) is known as an elementary stream. In transmission, many elementary streams will be combined to make a transport stream. Multiplexing requires blocks or packets of constant size. It is advantageous if these are short so that each elementary stream in the multiplex can receive regular data. A transport stream has a complex structure because it needs to incorporate metadata indicating which audio elementary streams and ancillary data are associated with which video elementary stream. It is possible to have a single program transport stream (SPTS) which carries only the elementary streams of one TV program.

For certain purposes, such as recording a single elementary stream, the transport stream is not appropriate. The small packets of the transport stream each require a header and this wastes storage space. In this case a program stream can be used. A program stream is a simplified

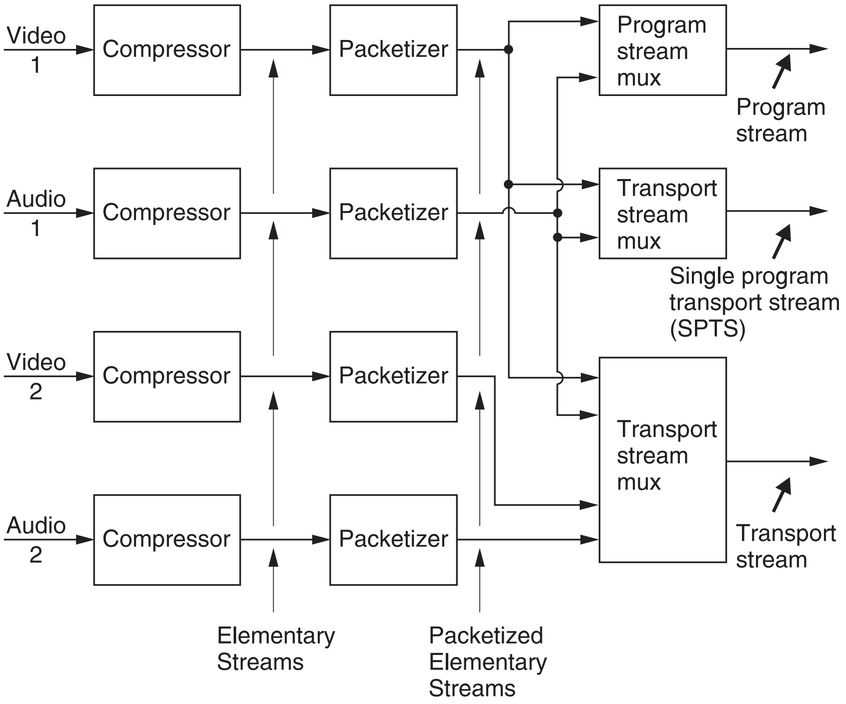


Figure 1.18 The bitstream types of MPEG-2. See text for details.

bitstream which multiplexes audio and video for a single program together, provided they have been encoded from a common locked clock. Unlike a transport stream, the blocks are larger and are not necessarily of fixed size.

1.14 Drawbacks of compression

By definition, compression removes redundancy from signals. Redundancy is, however, essential to making data resistant to errors. As a result, compressed data are more sensitive to errors than uncompressed data. Thus transmission systems using compressed data must incorporate more powerful error-correction strategies and avoid compression techniques which are notoriously sensitive. As an example, the Digital Betacam format uses relatively mild compression and yet requires 20 per cent redundancy whereas the D-5 format does not use compression and only requires 17 per cent redundancy even though it has a recording density 30 per cent higher. Techniques using tables such as the Lempel–Ziv–Welch codes are very sensitive to bit errors as an error in the transmission of a table value results in bit errors every time that table location is accessed. This is known as error propagation. Variable-length techniques such as the Huffman code are also sensitive to bit errors. As there is no fixed symbol size, the only way the decoder can parse a serial bitstream into symbols is to increase the assumed wordlength a bit at a time until a code value is recognized. The next bit must then be the first bit in the next symbol. A single bit in error could cause the length of a code to be wrongly assessed and then all subsequent codes would also be wrongly decoded until synchronization could be re-established. Later variable-length codes sacrifice some compression efficiency in order to offer better resynchronization properties.

In non-real-time systems such as computers an uncorrectable error results in reference to the back-up media. In real-time systems such as audio and video this is impossible and concealment must be used. However, concealment relies on redundancy and compression reduces the degree of redundancy. Media such as hard disks can be verified so that uncorrectable errors are virtually eliminated, but tape is prone to dropouts which will exceed the burst-correcting power of the replay system from time to time. For this reason the compression factors used on audio or video tape should be moderate.

As perceptive coders introduce noise, it will be clear that in a concatenated system the second codec could be confused by the noise due to the first. If the codecs are identical then each may well make, or better still be designed to make, the same decisions when they are in tandem. If the codecs are not identical the results could be disappointing. Signal manipulation between codecs can also result in artifacts which were previously undetectable becoming visible because the signal that was masking them is no longer present.

In general, compression should not be used for its own sake, but only where a genuine bandwidth or cost bottleneck exists. Even then the mildest compression possible should be used. Whilst high compression factors are permissible for final delivery of material to the consumer, they are not advisable prior to any post-production stages. For contribution material, lower compression factors are essential and this is sometimes referred to as *mezzanine* level compression.

One practical drawback of compression systems is that they are largely generic in structure and the same hardware can be operated at a variety of compression factors. Clearly the higher the compression factor, the cheaper the system will be to operate so there will be economic pressure to use high compression factors. Naturally the risk of artifacts is increased and so there is (or should be) counterpressure from those with engineering skills to moderate the compression. The way of the world at the time of writing is that the accountants have the upper hand. This was not a problem when there were fixed standards such as PAL and NTSC, as there was no alternative but to adhere to them. Today there is plenty of evidence that the variable compression factor control is being turned too far in the direction of economy.

It has been seen above that concatenation of compression systems should be avoided as this causes generation loss. Generation loss is worse if the codecs are different. Interlace is a legacy compression technique and if concatenated with MPEG, generation loss will be exaggerated. In theory and in practice better results are obtained in MPEG for the same bit rate if the input is progressively scanned. Consequently the use of interlace with MPEG coders cannot be recommended for new systems. Chapter 5 explores this theme in greater detail.

1.15 Compression pre-processing

Compression relies completely on identifying redundancy in the source material. Consequently anything which reduces that redundancy will have a damaging effect. Noise is particularly undesirable as it creates additional spatial frequencies in individual pictures as well as spurious differences between pictures. Where noisy source material is anticipated some form of noise reduction will be essential.

When high compression factors must be used to achieve a low bit rate, it is inevitable that the level of artifacts will rise. In order to contain the artifact level, it is necessary to restrict the source entropy prior to the coder. This may be done by spatial low-pass filtering to reduce the picture resolution, and may be combined with downsampling to reduce the number of pixels per picture. In some cases, such as teleconferencing, it will also be necessary to reduce the picture rate. At very low bit rates the use of interlace becomes acceptable as a pre-processing stage providing downsampling prior to the MPEG compression.

A compression pre-processor will combine various types of noise reduction (see Chapter 3) with spatial and temporal downsampling.

1.16 Some guidelines

Although compression techniques themselves are complex, there are some simple rules which can be used to avoid disappointment. Used wisely, MPEG compression has a number of advantages. Used in an inappropriate manner, disappointment is almost inevitable and the