1

Introduction

1.1 THE SCENE

Scene 1: Your avatar (a realistic 3D model with your appearance and voice) walks through a sophisticated virtual world populated by other avatars, product advertisements and video walls. On one virtual video screen is a news broadcast from your favourite channel; you want to see more about the current financial situation and so you interact with the broadcast and pull up the latest stock market figures. On another screen you call up a videoconference link with three friends. The video images of the other participants, neatly segmented from their backgrounds, are presented against yet another virtual backdrop.

Scene 2: Your new 3G vidphone rings; you flip the lid open and answer the call. The face of your friend appears on the screen and you greet each other. Each sees a small, clear image of the other on the phone's screen, without any of the obvious 'blockiness' of older-model video phones. After the call has ended, you call up a live video feed from a football match. The quality of the basic-rate stream isn't too great and you switch seamlessly to the higher-quality (but more expensive) 'premium' stream. For a brief moment the radio signal starts to break up but all you notice is a slight, temporary distortion in the video picture.

These two scenarios illustrate different visions of the next generation of multimedia applications. The first is a vision of MPEG-4 Visual: a rich, interactive on-line world bringing together synthetic, natural, video, image, 2D and 3D 'objects'. The second is a vision of H.264/AVC: highly efficient and reliable video communications, supporting two-way, 'streaming' and broadcast applications and robust to channel transmission problems. The two standards, each with their advantages and disadvantages and each with their supporters and critics, are contenders in the race to provide video compression for next-generation communication applications.

Turn on the television and surf through tens or hundreds of digital channels. Play your favourite movies on the DVD player and breathe a sigh of relief that you can throw out your antiquated VHS tapes. Tune in to a foreign TV news broadcast on the web (still just a postage-stamp video window but the choice and reliability of video streams is growing all the time). Chat to your friends and family by PC videophone. These activities are now commonplace and unremarkable, demonstrating that digital video is well on the way to becoming a ubiquitous

H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia.
Iain E. G. Richardson. ISBN: 0-470-84837-5

and essential component of the entertainment, computing, broadcasting and communications industries.

Pervasive, seamless, high-quality digital video has been the goal of companies, researchers and standards bodies over the last two decades. In some areas (for example broadcast television and consumer video storage), digital video has clearly captured the market, whilst in others (videoconferencing, video email, mobile video), market success is perhaps still too early to judge. However, there is no doubt that digital video is a globally important industry which will continue to pervade businesses, networks and homes. The continuous evolution of the digital video industry is being driven by commercial and technical forces. The commercial drive comes from the huge revenue potential of persuading consumers and businesses (a) to replace analogue technology and older digital technology with new, efficient, high-quality digital video products and (b) to adopt new communication and entertainment products that have been made possible by the move to digital video. The technical drive comes from continuing improvements in processing performance, the availability of higher-capacity storage and transmission mechanisms and research and development of video and image processing technology.

Getting digital video from its source (a camera or a stored clip) to its destination (a display) involves a chain of components or processes. Key to this chain are the processes of compression (encoding) and decompression (decoding), in which bandwidth-intensive 'raw' digital video is reduced to a manageable size for transmission or storage, then reconstructed for display. Getting the compression and decompression processes 'right' can give a significant technical and commercial edge to a product, by providing better image quality, greater reliability and/or more flexibility than competing solutions. There is therefore a keen interest in the continuing development and improvement of video compression and decompression methods and systems. The interested parties include entertainment, communication and broadcasting companies, software and hardware developers, researchers and holders of potentially lucrative patents on new compression algorithms.

The early successes in the digital video industry (notably broadcast digital television and DVD-Video) were underpinned by international standard ISO/IEC 13818 [1], popularly known as 'MPEG-2' (after the working group that developed the standard, the Moving Picture Experts Group). Anticipation of a need for better compression tools has led to the development of two further standards for video compression, known as ISO/IEC 14496 Part 2 ('MPEG-4 Visual') [2] and ITU-T Recommendation H.264/ISO/IEC 14496 Part 10 ('H.264') [3]. MPEG-4 Visual and H.264 share the same ancestry and some common features (they both draw on well-proven techniques from earlier standards) but have notably different visions, seeking to improve upon the older standards in different ways. The vision of MPEG-4 Visual is to move away from a restrictive reliance on rectangular video images and to provide an open, flexible framework for visual communications that uses the best features of efficient video compression and object-oriented processing. In contrast, H.264 has a more pragmatic vision, aiming to do what previous standards did (provide a mechanism for the compression of rectangular video images) but to do it in a more efficient, robust and practical way, supporting the types of applications that are becoming widespread in the marketplace (such as broadcast, storage and streaming).

At the present time there is a lively debate about which (if either) of these standards will come to dominate the market. MPEG-4 Visual is the more mature of the two new standards (its first Edition was published in 1999, whereas H.264 became an International

Standard/Recommendation in 2003). There is no doubt that H.264 can out-perform MPEG-4 Visual in compression efficiency but it does not have the older standard's bewildering flexibility. The licensing situation with regard to MPEG-4 Visual is clear (and not popular with some parts of the industry) but the cost of licensing H.264 remains to be agreed. This book is about these two important new standards and examines the background to the standards, the core concepts and technical details of each standard and the factors that will determine the answer to the question 'MPEG-4 Visual or H.264?'.

1.2 VIDEO COMPRESSION

Network bitrates continue to increase (dramatically in the local area and somewhat less so in the wider area), high bitrate connections to the home are commonplace and the storage capacity of hard disks, flash memories and optical media is greater than ever before. With the price per transmitted or stored bit continually falling, it is perhaps not immediately obvious why video compression is necessary (and why there is such a significant effort to make it better). Video compression has two important benefits. First, it makes it possible to use digital video in transmission and storage environments that would not support uncompressed ('raw') video. For example, current Internet throughput rates are insufficient to handle uncompressed video in real time (even at low frame rates and/or small frame size). A Digital Versatile Disk (DVD) can only store a few seconds of raw video at television-quality resolution and frame rate and so DVD-Video storage would not be practical without video and audio compression. Second, video compression enables more efficient use of transmission and storage resources. If a high bitrate transmission channel is available, then it is a more attractive proposition to send high-resolution compressed video or multiple compressed video channels than to send a single, low-resolution, uncompressed stream. Even with constant advances in storage and transmission capacity, compression is likely to be an essential component of multimedia services for many years to come.

An information-carrying signal may be compressed by removing redundancy from the signal. In a lossless compression system statistical redundancy is removed so that the original signal can be perfectly reconstructed at the receiver. Unfortunately, at the present time lossless methods can only achieve a modest amount of compression of image and video signals. Most practical video compression techniques are based on lossy compression, in which greater compression is achieved with the penalty that the decoded signal is not identical to the original. The goal of a video compression algorithm is to achieve efficient compression whilst minimising the distortion introduced by the compression process.

Video compression algorithms operate by removing redundancy in the temporal, spatial and/or frequency domains. Figure 1.1 shows an example of a single video frame. Within the highlighted regions, there is little variation in the content of the image and hence there is significant spatial redundancy. Figure 1.2 shows the same frame after the background region has been low-pass filtered (smoothed), removing some of the higher-frequency content. The human eye and brain (Human Visual System) are more sensitive to lower frequencies and so the image is still recognisable despite the fact that much of the 'information' has been removed. Figure 1.3 shows the next frame in the video sequence. The sequence was captured from a camera at 25 frames per second and so there is little change between the two frames in the short interval of $1/25$ of a second. There is clearly significant temporal redundancy, i.e. most

Figure 1.1 Video frame (showing examples of homogeneous regions)

Figure 1.2 Video frame (low-pass filtered background)

Figure 1.3 Video frame 2

of the image remains unchanged between successive frames. By removing different types of redundancy (spatial, frequency and/or temporal) it is possible to compress the data significantly at the expense of a certain amount of information loss (distortion). Further compression can be achieved by encoding the processed data using an entropy coding scheme such as Huffman coding or Arithmetic coding.

Image and video compression has been a very active field of research and development for over 20 years and many different systems and algorithms for compression and decompression have been proposed and developed. In order to encourage interworking, competition and increased choice, it has been necessary to define standard methods of compression encoding and decoding to allow products from different manufacturers to communicate effectively. This has led to the development of a number of key International Standards for image and video compression, including the JPEG, MPEG and H.26× series of standards.

1.3 MPEG-4 AND H.264

MPEG-4 Visual and H.264 (also known as Advanced Video Coding) are standards for the coded representation of visual information. Each standard is a document that primarily defines two things, a coded representation (or syntax) that describes visual data in a compressed form and a method of decoding the syntax to reconstruct visual information. Each standard aims to ensure that compliant encoders and decoders can successfully interwork with each other, whilst allowing manufacturers the freedom to develop competitive and innovative products. The standards specifically do not define an encoder; rather, they define the output that an encoder should

produce. A decoding method is defined in each standard but manufacturers are free to develop alternative decoders as long as they achieve the same result as the method in the standard.

MPEG-4 Visual (Part 2 of the MPEG-4 group of standards) was developed by the Moving Picture Experts Group (MPEG), a working group of the International Organisation for Standardisation (ISO). This group of several hundred technical experts (drawn from industry and research organisations) meet at 2–3 month intervals to develop the MPEG series of standards. MPEG-4 (a multi-part standard covering audio coding, systems issues and related aspects of audio/visual communication) was first conceived in 1993 and Part 2 was standardised in 1999. The H.264 standardisation effort was initiated by the Video Coding Experts Group (VCEG), a working group of the International Telecommunication Union (ITU-T) that operates in a similar way to MPEG and has been responsible for a series of visual telecommunication standards. The final stages of developing the H.264 standard have been carried out by the Joint Video Team, a collaborative effort of both VCEG and MPEG, making it possible to publish the final standard under the joint auspices of ISO/IEC (as MPEG-4 Part 10) and ITU-T (as Recommendation H.264) in 2003.

MPEG-4 Visual and H.264 have related but significantly different visions. Both are concerned with compression of visual data but MPEG-4 Visual emphasises flexibility whilst H.264's emphasis is on efficiency and reliability. MPEG-4 Visual provides a highly flexible toolkit of coding techniques and resources, making it possible to deal with a wide range of types of visual data including rectangular frames ('traditional' video material), video objects (arbitrary-shaped regions of a visual scene), still images and hybrids of natural (real-world) and synthetic (computer-generated) visual information. MPEG-4 Visual provides its functionality through a set of coding tools, organised into 'profiles', recommended groupings of tools suitable for certain applications. Classes of profiles include 'simple' profiles (coding of rectangular video frames), object-based profiles (coding of arbitrary-shaped visual objects), still texture profiles (coding of still images or 'texture'), scalable profiles (coding at multiple resolutions or quality levels) and studio profiles (coding for high-quality studio applications).

In contrast with the highly flexible approach of MPEG-4 Visual, H.264 concentrates specifically on efficient compression of video frames. Key features of the standard include compression efficiency (providing significantly better compression than any previous standard), transmission efficiency (with a number of built-in features to support reliable, robust transmission over a range of channels and networks) and a focus on popular applications of video compression. Only three profiles are currently supported (in contrast to nearly 20 in MPEG-4 Visual), each targeted at a class of popular video communication applications. The Baseline profile may be particularly useful for "conversational" applications such as videoconferencing, the Extended profile adds extra tools that are likely to be useful for video streaming across networks and the Main profile includes tools that may be suitable for consumer applications such as video broadcast and storage.

1.4 THIS BOOK

The aim of this book is to provide a technically-oriented guide to the MPEG-4 Visual and H.264/AVC standards, with an emphasis on practical issues. Other works cover the details of the other parts of the MPEG-4 standard [4–6] and this book concentrates on the application of MPEG-4 Visual and H.264 to the coding of natural video. Most practical applications of

MPEG-4 (and emerging applications of H.264) make use of a subset of the tools provided by each standard (a 'profile') and so the treatment of each standard in this book is organised according to profile, starting with the most basic profiles and then introducing the extra tools supported by more advanced profiles.

Chapters 2 and 3 cover essential background material that is required for an understanding of both MPEG-4 Visual and H.264. Chapter 2 introduces the basic concepts of digital video including capture and representation of video in digital form, colour-spaces, formats and quality measurement. Chapter 3 covers the fundamentals of video compression, concentrating on aspects of the compression process that are common to both standards and introducing the transform-based CODEC 'model' that is at the heart of all of the major video coding standards.

Chapter 4 looks at the standards themselves and examines the way that the standards have been shaped and developed, discussing the composition and procedures of the VCEG and MPEG standardisation groups. The chapter summarises the content of the standards and gives practical advice on how to approach and interpret the standards and ensure conformance. Related image and video coding standards are briefly discussed.

Chapters 5 and 6 focus on the technical features of MPEG-4 Visual and H.264. The approach is based on the structure of the Profiles of each standard (important conformance points for CODEC developers). The Simple Profile (and related Profiles) have shown themselves to be by far the most popular features of MPEG-4 Visual to date and so Chapter 5 concentrates first on the compression tools supported by these Profiles, followed by the remaining (less commercially popular) Profiles supporting coding of video objects, still texture, scalable objects and so on. Because this book is primarily about compression of natural (real-world) video information, MPEG-4 Visual's synthetic visual tools are covered only briefly. H.264's Baseline Profile is covered first in Chapter 6, followed by the extra tools included in the Main and Extended Profiles. Chapters 5 and 6 make extensive reference back to Chapter 3 (Video Coding Concepts). H.264 is dealt with in greater technical detail than MPEG-4 Visual because of the limited availability of reference material on the newer standard.

Practical issues related to the design and performance of video CODECs are discussed in Chapter 7. The design requirements of each of the main functional modules required in a practical encoder or decoder are addressed, from motion estimation through to entropy coding. The chapter examines interface requirements and practical approaches to pre- and post-processing of video to improve compression efficiency and/or visual quality. The compression and computational performance of the two standards is compared and rate control (matching the encoder output to practical transmission or storage mechanisms) and issues faced in transporting and storing of compressed video are discussed.

Chapter 8 examines the requirements of some current and emerging applications, lists some currently-available CODECs and implementation platforms and discusses the important implications of commercial factors such as patent licenses. Finally, some predictions are made about the next steps in the standardisation process and emerging research issues that may influence the development of future video coding standards.

1.5 REFERENCES

1. ISO/IEC 13818, Information Technology – Generic Coding of Moving Pictures and Associated Audio Information, 2000.

2. ISO/IEC 14496-2, Coding of Audio-Visual Objects – Part 2:Visual, 2001.
3. ISO/IEC 14496-10 and ITU-T Rec. H.264, Advanced Video Coding, 2003.
4. F. Pereira and T. Ebrahimi (eds), *The MPEG-4 Book*, IMSC Press, 2002.
5. A. Walsh and M. Bourges-Sévenier (eds), *MPEG-4 Jump Start*, Prentice-Hall, 2002.
6. ISO/IEC JTC1/SC29/WG11 N4668, MPEG-4 Overview, http://www.m4if.org/resources/Overview.pdf, March 2002.