MPEG-4 and the Challenges of Benchmarking its Implementations for Picture Quality

A. Punchihewa, D.G. Bailey, and R. M. Hodgson
Institute of Information Sciences & Technology, Massey University, Palmerston North, New Zealand
G.A.Punchihewa@massey.ac.nz

Abstract

An important aspect of image and video coding is quality assessment. The MPEG family uses many techniques to compress image and video data resulting visual artefacts. ISO 14496 standardises the implementation of MPEG-4 codec, by specifying the bit stream syntax. In addition to the techniques used by MPEG-1 and MPEG-2, MPEG-4 adds object coding, enabling different compression techniques to be used for different objects within the image. The objective measures used for quality assessment in MPEG-1 and 2 may also be used on those components in MPEG-4. However new techniques must be developed to exercise and assess the new features of MPEG-4. New artefacts introduced by MPEG-4 are those related to segmentation of objects from background, and texture scaling artefacts as textures are mapped onto moving shapes. The major challenge to benchmarking picture quality is that the standard only specifies the bit stream and decoding process, allowing a wide range of coders to be implemented for the same video sequence. As quality is determined primarily by the coding process different qualities may be achieved by different implementations even for the same input data, and the same output bitrate.

Keywords: image, video, quality, artefacts, subjective, objective, coding, MPEG-4

1 Introduction

Image and video communication technologies have been advancing at a rapid pace for the last two decades. For most digital video codecs, increasing bit rate performance has been achieved at the cost of increasing complexity in techniques and implementations. McCann [3] claims that codec development achieves an average of a 15% reduction in bit rate per annum. Consumer products that are available today include a broad range of new imaging applications. JPEG and MPEG codecs are in common use in digital photography, digital TV, photographic quality printers and DVDs. Most of these use block-transform based techniques and consequently produce visually annoying compression artefacts. World Wide Web (WWW), television and multimedia networks also deploy numerous codecs. Providers and end users are using codecs to transfer content between production facilities and to deliver, to interact with and to manage content. However there is a problem related to picture quality, especially from the television engineers' perspective, where the picture quality is well below conventional broadcast pictures due to errors in colour, coding and optics. Although picture quality evaluation is a complex subject that needs to be performed to ensure an agreed level of service from a professional broadcasters' point of view [4].

2 What is MPEG-4?

MPEG-4 is the most recently emerged digital multimedia standard with associated protocols for representing, manipulating and transporting natural and synthetic multimedia content (i.e. audio, video and data) over a broad range of communication infrastructures. The MPEG-2 system has been in use since the mid 1990s. MPEG-4, the subsequent generation system, is an ISO/IEC standard developed by MPEG (Moving Picture Experts Group), the committee that also developed the MPEG-1 and MPEG-2. These standards made interactive video on CD-ROM and Digital Television possible.

The process of international standardization takes place slowly, and to ensure that a standard is eventually achieved there are strict rules that prohibit substantive change after a certain point in the process. So, by the time a standard is officially adopted there is often a backlog of desired enhancements and extensions. So it was with MPEG-2. During this time period, MPEG-3 had been started and abandoned, so the next project became MPEG-4. Two versions of MPEG-4 are already complete, and work is continuing on further extensions. At first, the main focus of MPEG-4 was the encoding of video and audio at very low rates. In fact, the standard was explicitly optimized for three bit rate ranges: below 64 kbits/s, 64 - 384 kbits/s, and 384 kbits/s - 4 Mbits/s. Better performances at low bit rates remained a major objective and some very creative ideas contributed to this end. Great attention was also paid to error resilience, making MPEG-4 very suitable for use in error-prone environments, such as transmission to personal handheld devices. However, other profiles and levels use bit rates up to 38.4 Mbits/s, and work is still proceeding on studio-quality profiles and levels using data rates up to 1.2 Gbits/s. In transmission, MPEG-4 data has built-in error resilience through forward error correction capability and can be used for QoS management when offering video over IP networks. MPEG-4 audio visual objects (AVO) are classified based on application-level Quality of Service (QoS) criteria.

MPEG-4 is the result of another international effort involving hundreds of researchers and engineers from all over the world. MPEG-4, or ISO/IEC 14496, was finalized in October 1998 and became an International Standard in the first months of 1999. The fully backward compatible extensions under the title of MPEG-4 Version 2 were frozen at the end of 1999, to acquire the formal International Standard Status early in 2000. The MPEG-4 visual standard completed almost 5 years ago continues to be improved. MPEG-4 Part 4, included not just more advanced compression tools but also a potential new way of delivering multimedia with object and semantic coding. However, the part which has been most used is the one dealing with new compression tools. The MPEG-4 standard now (part 10) includes an advanced video codec jointly developed between Video Coding Expert Group (VCEG) of International Telecommunication Union (ITU) and the MPEG committee of ISO/IEC. It is known as "Advanced Video Coding" (AVC) and is designated by H.264 by ITU [4, 5, 6, 7]. Like the other MPEG standards, MPEG-4 specifies the decoding process and the bit stream syntax. This leads to many encoder and decoder implementations. The MPEG-4 development team claims that MPEG-4 builds on the proven success of three fields: digital television, interactive graphics applications (synthetic content) and interactive multimedia in form of World Wide Web for distribution of and access to content. MPEG-4 provides the standardized technological elements enabling the integration of the production, distribution and content access in the above three fields. More importantly, MPEG-4 has became more than just another compression system as it evolved into a comprehensive system for multi-media encoding with complete powerful tools for interactivity and a large range of applications [3].

2.1 Scope of MPEG-4

The content; video, audio and meta-data[1] are handled by three major groups namely content creators (providers), content delivery providers (network service providers) and consumers of content. For content creators, MPEG-4 enables the production of content that has far greater reusability, has greater flexibility than is possible today with individual technologies such as digital television, animated graphics, World Wide Web (WWW) pages and their extensions. Also, it is now possible to better manage and protect content ownership rights. For end users, MPEG-4 brings higher levels of interaction with content, within the limits set by the content author. It also brings multimedia to new networks, including those employing relatively low bit rate, and mobile networks. AVC gains 40-50% compression efficiency over today's optimized MPEG-2 bit rates for television broadcasters [4, 5].

MPEG-4 is broad in its scope, and the descriptions above only cover a few of the many aspects of the standard. There are studio profiles for high-quality encoding which, in conjunction with object coding will permit structured storage of all the separate elements of a video composite. Facial and body animation profiles will permit a stored face to "read" text in many languages. Further extensions of MPEG-4 may even provide solutions for digital cinema. Different applications use a different profile or subset of MPEG-4, designed for that application. The lowest level of profile MPEG-4 is the baseline profile, which is used in videophone. An extended profile is used for streaming and main profiles are provided for broadcast and interlaced video. Different profiles have a limited number of object handling capabilities and related complexity. An estimated two fold improvement in coding efficiency comes at the cost of a four fold increase in decoder complexity when compared with MPEG-2 [5].

2.2 Coding in MPEG-4

MPEG-4 achieves the aims described in section 2.1 by providing standardized ways to represent units of aural, visual or audiovisual content, called "media objects". MPEG-4 audiovisual scenes are composed of multiple media objects that are organized in a hierarchical structure. Primitive media objects include: still images (e.g. as a fixed background); video objects (e.g. a talking person - without the background; audio objects (e.g. the voice associated with that person, background music) in the hierarchy. MPEG-4 standardizes a number of such primitive media objects, capable of representing both natural and synthetic content types, which can be either 2- or 3-dimensional. These objects can be recorded with a camera or microphone, or generated with a computer. MPEG-4 provides a standard way to describe the composition of these objects to create compound media objects that form audiovisual scenes; multiplex and synchronize the data associated with media objects, so that they can be transported over network channels providing a QoS appropriate for the nature of the specific media objects; and interact with the audiovisual scene generated at the receiver's end [7]. The complete scene is segmented to foreground and background. Background is texture encoded and

---

[1] Meta-data are data about data that describes how, when and by whom a particular set of data was collected, and how the data is formatted. MPEG-7 provides a standardised method to present the.

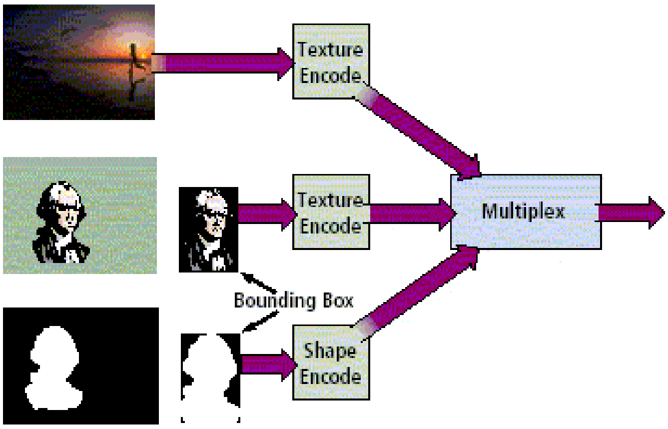

Figure 1: Object Coding -Textures and Shapes [8]

foreground object is split into shape and the texture, which are encoded separately. Then all three components are multiplexed to a single bit stream as illustrated in Figure 1.

2.3 Object Coding

The concept of objects is the most significant difference from conventional transmission systems. Different parts of the input scene can be coded and transmitted separately as video objects and audio objects to be brought together, or composited, by the decoder. Different object types may each be coded with the tools most appropriate to the job. The objects may be generated independently, or a scene may be analyzed to separate objects. For example, the foreground and background objects in video coverage of a soccer game can be processed to separate the ball from the rest of the scene. The background (the scene without the ball) can be transmitted separately to attract a pay-per-view audience. Anyone could see the players and the field, but only those who pay can see the ball![8]. The object-oriented approach leads to three key characteristics of MPEG-4 streams: (i) Multiple objects may be encoded using different techniques, and composited at the decoder, (ii) objects may be of natural origin, such as scenes from a camera, or synthetic, such as text, and (iii) instructions in the bit stream or user choice may enable several different presentations to be composed from the same bit stream.

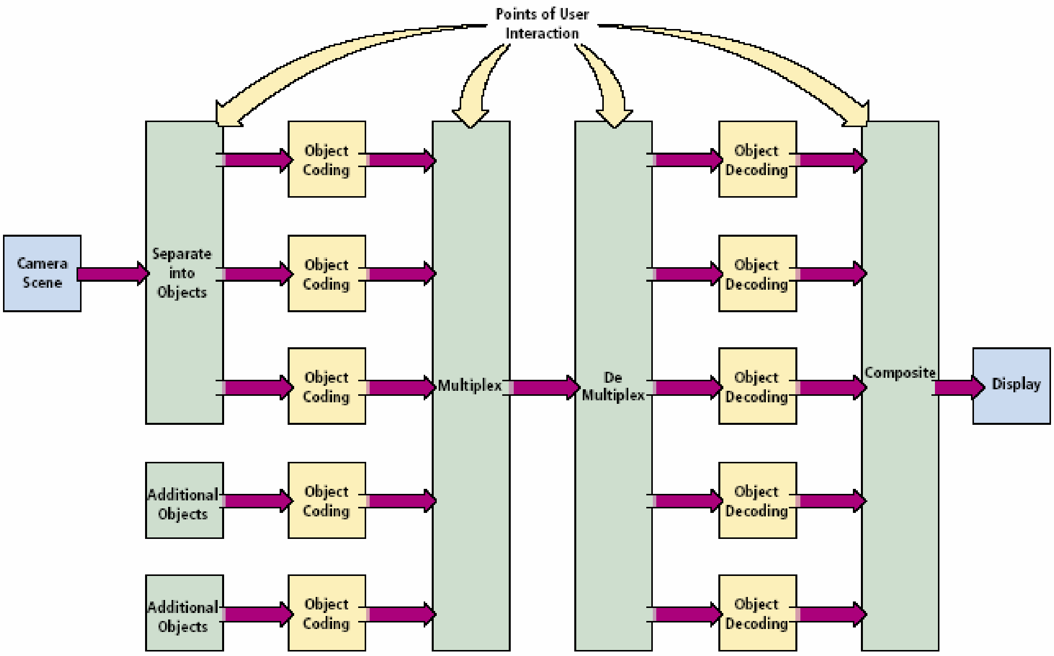

Figure 2: Object Coding in MPEG-4 [8]

The generalized system for object coding in MPEG-4 is shown in Figure 2. This diagram also emphasizes the opportunities for user interaction within MPEG-4 systems which is a powerful feature, particularly for video game designers. These capabilities do not have to be used as MPEG-4 provides traditional coding of video and audio, and improvements on MPEG-2 by offering better coding efficiency and resilience to errors. However, the true potential of MPEG-4 comes from the architecture described above. The independent coding of objects offers a number of advantages. Each object may be coded in the most efficient manner, and different spatial or temporal scaling may be used as appropriate.

Many of the video coding tools in MPEG-4 are similar to those of MPEG-2, but enhanced by better use of predictive coding and more efficient entropy coding. In the simplest model, video is coded in much the same way as in MPEG-2, but it is described as a single video object with a rectangular shape. The representation of the image is known as texture coding. Where there is more than one video object, some may have irregular shapes, and generally all will be smaller than a full screen background object. This means that only the active area of each object needs be coded, but the shape and position must also be represented. The standard includes tools for shape coding of rectangular and irregular objects, in either binary or grey-scale representations. The concept is shown in Figure 1.

MPEG-4 offers PSNR quality scalability. In the context of media compression, scalability means the ability to distribute content at more than one quality level within the same bit stream. MPEG-2 and MPEG-4 both provide scalable profiles using a conventional model: the encoder generates a base-layer and one or more enhancement layers. Scalability is based on refining the amplitude resolution. The enhancement layer(s) may be discarded for transmission or decoding if insufficient resources are available. The concept is illustrated in the Figure 3. This approach works, but all decisions about quality levels have to be made at the time of encoding, and in practice the number of enhancement layers is severely limited (usually to one).

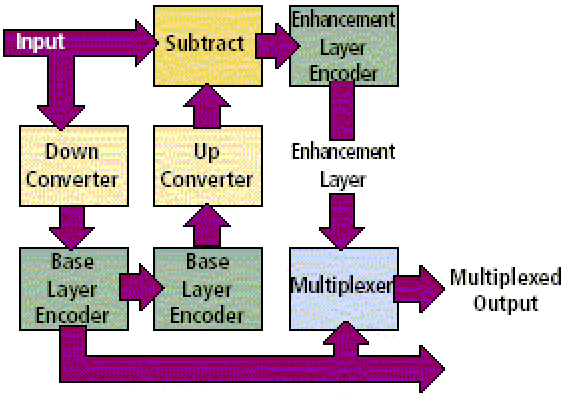

Figure 3: Concept of a Spatially-Scaled Encoder [8]

3 MPEG-4 Picture Quality and Artefacts

Based on the conventional digital communication theories, there are two main sources of artifacts. They are (i) the source coding where coding is done for efficient representation of information and (ii) transmission process. Source coding digital images to exploit the redundancy present inevitably results in a loss of quality in the resultant image or video. This is because in removing redundant information, often the some of the essential information is distorted or destroyed, resulting in artefacts, impairments or defects [1, 2]. Transmission errors are mitigated by channel coding where redundancy bits are added to the information bits, for error detection and correction at the receive end. Artefacts create undesirable visual effects in the picture resulting from technical limitations of the coding process. The compression algorithms used, the picture content, and the origin of the source material contribute to the coding artefacts. The greatest technical limitation is the available bandwidth, which affects the compression ratio and data rate. In general, artefacts will become more visible as the compression ratio is increased. Compression-related artefacts such as blockiness, blur and ringing are common at low bit rates. Blocking artefacts are the result of the independent processing of each block in block-based signal processing. Blur and ringing result from truncation or quantisation of coefficients in the frequency domain. Ringing artefacts typically appear as sharp oscillations along the edges of an object against a relatively uniform background. Any motion of the object in a video results in these oscillations flickering, giving mosquito noise [1, 2]. At low bit rates, PSNR is unable to assess the visual quality of image sequences, moreover PSNR deals with image fidelity, and not with image quality. Video quality differs from still image quality as video is a sequence of images, which may contain intra frame object movements. Errors in motion estimation and scan-conversions can result in jerkiness artefacts.

Compression technologies are evolving in quick succession, followed by image products for the professional and consumer markets. Television engineers need to evaluate picture quality for the services that television broadcasters offer in particular broadcast systems and components are usually evaluated for acceptable picture quality using known test signals and known measurement techniques. Examples of test signals used in analogue colour television broadcasting are colour bars, multi-burst, 20T pulse and T pulse which are used to measure hue errors, cross talk, inter-modulation, frequency response, non-linearity, system response and ghosting with objective measurements. Since coding technologies can be used in many configurations, special test signals and accompanying quality metrics can be used to benchmark different codecs or implementations. The technology adopters can make use of objective picture quality metrics in making prudent decisions when selecting a codec or codec standard as the lifetime of specific system is often very short in comparison to the development time. Greater complexity means more processing power is needed for codec, hence a higher cost resulting in trade offs in codec implementation.

4 Benchmarking Issues

Benchmarking is a continuous process of evaluating objective quality metrics to identify picture quality of a codec compared to another. The challenges of benchmarking MPEG-4 are due to the rapid pace of development, increased complexity of the MPEG-4 standard itself. Though many of the objective quality metrics developed by other researchers for digital image or video have been developed to work with natural scenes, images or video, known referenced (synthetic) objective picture quality assessment, generally offers swift, economical and uniform evaluation provided that the processing hierarchy is well represented [10]. The main goal in using such approach to benchmark a codec is to design a synthetic picture or sequence which would highlight the artefacts. The contents are usually designed to stress the codec and to evaluate the picture quality of the reconstructed picture. MPEG-4 can handle synthetic objects with shape coding and may handle them well. This would offer a challenge in designing the shape of a synthetic object that is effective in evaluating the new coding process.

Some of the compression technologies MPEG-4 uses include DCT, wavelets, motion estimation, object coding and error resilient coding. MPEG-4 only standardises the decoding process and the bit stream syntax, and this leads to many possible encoder and decoder implementations. This makes the quality metric design more complex as we need to evaluate many techniques, architectures and combinations of them. The same image source may be encoded in many different ways, with each approach using different object hierarchies and compression technologies, and yet all conform to the MPEG-4 standard. This complicates design of test images and sequences which stress particular components because the codec may not use those components.

Segmentation of video into objects is a function of the encoder and is not standardized. Synthetically generated content may already be segmented. In studio environment, techniques such as blue-screening may be used. Segmentation of video into objects at the capture end may introduce artifacts such as a borderline around the foreground object.

Since DCT is a block processing technique, we could anticipate block artefacts. In MPEG-4 artifacts are mitigated by deblocking and deringing filters, which are included in the post-processing of MPEG-4 decoder to improve the picture quality. MPEG-4 has the provision to use deringing filters. Ringing noise can be minimised by pre-processing the picture data where processing power is less constrained at the source of data [10]. Arbitrary shape coding in MPEG-4 may lead to aliasing when the shapes are coded as binary data, which is the simplest approach and indicate whether the object is transparent or opaque at any given point. The edge between the object and the background is abrupt. This leads to a violation of Nyquist theorem, and aliasing is likely to be visible.

In transmission, the objective quality metrics may represent application-level QoS levels where QoS parameters may include loss of blocks and jitter to reflect the impact on image or video quality. The QoS metric is based on the probability of lost blocks can be used to assess performance of a codec transportation module implementation. A new challenge with MPEG-4 would be to evaluate the performance of built-in error resilience on IP and other lossy packet networks.

Multimedia content viewed on smaller screens with progressive scan display may lead to display artifacts such as judder due to interlace to progressive scan conversion. This is usually most noticeable in sequences containing motion. MPEG-4 has the ability to map a static texture onto a varying shape and this is performed at the decoder. The varying shape and the static texture are transmitted as separate objects. The movement of the shape may cause stretching or expansion of the texture far beyond its normal size. Due to poor scaling capability, this can result in artefacts on the final display. Clearly static-texture-scaling artefacts are content dependent.

5 Summary

MPEG-4 offers greater flexibility in implementation than previous standards but it also accompany many challenges in benchmarking its implementations for picture quality, due to rapid improvements, increasing complexity, variations in implementations, and segmentation. Error resilience for transmission is a useful feature in transmission over IP networks. Texture coding and shape coding will introduce new forms of artifacts in MPEG-4. The major challenge to benchmarking picture quality of MPEG-4 is that the standard only specifies the bit stream and decoding process allowing wide range of mature and emerging compression technologies.

5 Acknowledgments

We wish to acknowledge the material adopted from previous publications and to express our gratitude to those authors.

6 References

[1] A. Punchihewa and D. G. Bailey: Artefacts in Image and Video Systems; Classification and Mitigation, Proceedings of Image and Vision Computing New Zealand 2002, pp 197-202 (2002).

[2] A. Punchihewa, D. G. Bailey: and R. M. Hodgson: A Survey of Coded Image and Video Quality Assessment, Proceedings of Image and Vision Computing New Zealand 2003, pp 326-331 (2003).

[3] Centre National Research Technologies in France, "Multimedia page", http://tim.irisa.fr/veille/TVadsl/Broadband%20TV/Broadband%20TVSEM/Presentations/Day%202%20morning/MPEG-4%20Video/H264_broadbandTV.pdf (August 2004).

[4] D. Wood: Everything you want to know about Video Codecs, EBU Technical review, European Broadcasting Union, Geneva (July 2003).

[5] MPEG Industry forum, "Home Page", http://www.m4if.org/mpeg4/, (August 2004).

[6] ISO website: www.iso.ch/ittf (August 2004).

[7] Koenen, R., Overview of the MPEG-4 Standard, ISO/IEC JTC1/SC29/WG11 Coding of moving pictures and audio, (March 2002).

[8] Broadcast papers, "Editing Page”,

http://www.broadcastpapers.com/editing/GVGTheNewMPEGs07.htm, (August 2004).

[9] P. D. Symes: Video compression demystified, PP194-219,McGraw Hill Publication 2001.

[10] Tektronix “video test page”, www.tektronix.com (August 2004).