New Trends in Multimedia Standards: MPEG4 and JPEG2000

Jie Liang
Media Technology Lab, DSPS R&D Center, Texas Instruments

liang@ti.com

Abstract

The dramatic increase in both computational power, brought on by the introduction of increasingly powerful chips, and the communications bandwidth, unleashed by the introduction of cable modem and ADSL, lays a solid foundation for the take-off of multimedia applications. Standards always play an important role in multimedia applications due to the need for wide distribution of multimedia contents. Standards have long played pivotal roles in the development of multimedia equipment and contents. MPEG4 and JPEG2000 are two recent multimedia standards under development under the auspice of the International Standards Organization (ISO). These new standards introduce new technology and new features that will become enabling technology for many emerging applications. In this paper, we describe the new trends and new developments that shape these new standards, and illustrate the potential impact of these new standards on multimedia applications.

Keywords: image coding, video coding, MPEG4, JPEG2000, multimedia.

Introduction

Multimedia has long played an important role in the process of informing activities: learning, studying, researching, and communicating, which can be summed up in an old Chinese saying, "*A picture is worth one thousand words*". The ability to see, hear, and interact with the material that a client is interested in provides valuable additional dimension to a client's learning experience. With the introduction of personal computers in 1980s, CD-ROM based learning has long been an established part of the educational process and a fast developing market. However, with the fast adoption of the Internet and the continuous development in PC processor speed and capability, a promising new era in multimedia, both in informing and entertainment applications, is just beginning. With an alarming rate, much of the world's knowledge base is moving on-line. It is already a reality in many parts of the United States that school children can access a web site over a cable modem or ADSL modem at home, pulling off a detailed marketing report for his/her term paper. In the near future, he would also be able to listen to a live broadcast of a lecture on the same subject he is researching, interact with the materials through an animated graphical user interface, and watch a short clip of video on that subject, all through a single standardized file format and protocol.

Standards have long played pivotal roles in the development of multimedia equipment and contents. The need for standardization in multimedia applications is not a choice, but a necessity. To be able to widely distribute multimedia content to readers, we must be sure that the format of the files that contain the media content can be recognized and decoded correctly no matter what devices or platforms the users are using. On the other hand, standards have significantly facilitated the spread of multimedia contents. Just look at the impact of JPEG on the distribution of images, and MPEG1/MPEG2 on the distribution of video.

In recent years, responding to the development of many promising new technologies in the field of multimedia and the practical needs of many commercial applications, the Geneva-based International Standards Organization (ISO) has undertaken two new projects in the last few years: MPEG4 and JPEG2000. MPEG4 is aimed at standardizing a framework for the representation and delivery of multimedia contents such as video, audio, text, and graphics in a unified framework, while JPEG2000 is focused on developing the next-generation image compression and representation standard. It is widely expected that the introduction of these two standards will become a significant catalyst driving the future growth of the multimedia industry, and will provide better tools for all participants in the informing science areas.

In the following, we introduce in more details the new technologies in the upcoming MPEG4 and JEPG2000 standards. Section 2 covers MPEG4, and JPEG2000 is covered in Section 3. We conclude by summarizing the new technology and

features, and elaborate more on their implications for multi-media applications.

The MPEG4 Video and Multimedia Standard

MPEG4 Overview

MPEG-4 is an ISO/IEC standard being developed by MPEG (Moving Pictures Experts Group). MPEG also developed the Emmy award winning standards MPEG-1 and MPEG-2. MPEG-1 was an audiovisual coding standard aimed at addressing the storage and retrieval of multimedia information on a CD-ROM [1]. MPEG-2 followed closely behind MPEG-1 and was a standard that addressed the broadcast TV applications. MPEG-2 was a hugely successful standard with significant acceptance in the market place with a number of other applications in addition to broadcast TV [1]. Some of the prominent applications of MPEG-2 include Direct Broadcast Satellite (DBS), Digital Versatile Disk (DVD) and High Resolution Television (HDTV). Initially MPEG-3 was reserved for HDTV applications. However, later on MPEG-2 was found to be suitable for HDTV, and it was decided to include HDTV as a separate profile of MPEG-2 and discontinue the MPEG-3 work item.

Both MPEG-1 and MPEG-2 use block-based motion compensation for removing time-domain redundancy, and use DCT-based block coders for encoding the residue and Intra-coded frames. Intra-coded frames are the frames that are coded with no references to other frames, and the whole frame is coded as a still image. In video sequences, adjacent frames have much overlapped information and a lot of redundancy can be removed by using motion compensation. In MPEG, a macro-block (16x16) motion compensation method is adopted, in which a macro-block in the current frame is matched to a macro-block in the reference frame, and only residue signal needs to be coded. For residue coding, DCT transform is used to convert the signal to the frequency domain, which represents better form for compression. At the system level, the compressed video bitstream is mixed and synchronized with audio signal so that complete movie or TV content can be delivered.

MPEG-4 is the next audio-visual coding standard from ISO after MPEG-1 and MPEG-2. Unlike the previous two standards, which had one clear application in mind when they were developed, MPEG-4 is a much broader umbrella type standard that has a number of different technologies, which are targeted at different applications. Initially, MPEG-4 was aimed primarily at low bit rate video communications. However, its scope was later expanded for it to be much more of a multimedia coding standard. MPEG-4 is efficient across a wide variety of bitrates ranging from a few Kbits/sec to tens of Mbits/sec. In addition to providing improved coding efficiency, MPEG-4 also provides a number of functionality. These include, 1) the ability to efficiently encode mixed media data such as video, graphics, text, images, audio and speech, (called as audio-visual objects AVOs), 2) the ability to create a compelling multi-media presentation by compositing these mixed media objects by a compositing script, 3) error resilience to enable robust transmission of compressed data over noisy communication channels, 4) the ability to encode arbitrary shaped video objects, 5) multiplex and synchronize the data associated with these objects, so that they can be transported over network channels providing a QoS appropriate for the nature of the specific objects and, 6) the ability to interact with the audio visual scene generated at the receiver end. These functionalities supported by the standard will enable many compelling applications ranging from wireless videophones, Internet multimedia presentations, broadcast TV and DVD. As can be imagined a standard that supports these diverse set of functionalities and associated applications is fairly complex. This paper focuses only on some major innovations of the MPEG4 video coder. The reader is directed to the wealth of information that is available on MPEG at the official MPEG web site http://www.cselt.it/mpeg and other recent MPEG4 publications [3,4,5,6,7,8].

MPEG-4 reached a Committee Draft (CD) standard status in November 1997. MPEG-4 was released in October 1998 as Draft International Standard (DIS), and will be an official International Standard (IS) in 1999. MPEG-4 Committee has also taken the approach of "versioning" to the standard formation process. MPEG-4 version 1 includes a number of useful tools and version 2 is expected to include some others that are being developed by the standards body [3]. It is expected that MPEG-4 version 2 will be backwards compatible with MPEG-4 version 1.

Being a multimedia coding standard, MPEG-4 standardizes tools not only for video coding but also for coding audio, graphics and text. The standard also includes a systems part, which describes how the audio, video, text and graphics are synchronized and presented in a compelling manner for various applications. Next, we highlight some of the most novel parts in the emerging MPEG4 standards.

Figure 1. An MPEG4 scene with audio and video objects

Object-base Multimedia Composition and Interactivity

The basic building block of a multimedia presentation in MPEG4 is *object*. There are many basic object types supported in

MPEG4: video objects, audio objects, texts, graphics, 2-D meshes, 3-D meshes, HTML objects, etc. These objects are placed inside a scene with descriptors indicating its location, orientation, and other necessary properties. A scene in MPEG4 is described using *Binary Format for Scenes (BIFS)* – an efficient syntax for describing the compositions of a scene with binary data. An example of a MPEG4 scene is given in Figure 1. In this scene, we have a moving video ob-

ject – the teacher. An audio object is also associated with the video object that stores the speeches spoken by the teacher. We also have graphic objects such as the globe, the desk, and the white board. Simultaneously, a dynamic web page object is displayed on the white board – the presentation material for the course. Each object has a built-in timing attribute, which can be utilized to synchronize the presentation timing of different objects. For instance, the synchronization of the lip movement of the video object and the audio object, and the synchronization of the audio object and the page changes of the HTML object on the white board.

The objects and the BIFS define the actual contents and the spatial and temporal relationship among the different objects
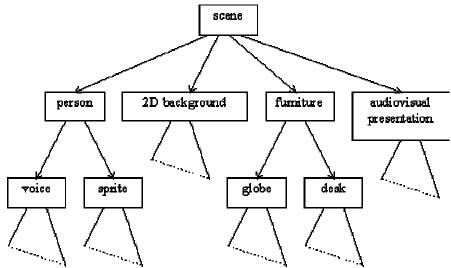

Figure 2. The Tree Structure of An MPEG4 Scene

inside a scene. An MPEG4 compositor performs the actual presentation of the scene. The role of a compositor is to take the decoder output of each object, look at the attribute fields of the object descriptor and the information contained in the BIFS stream, and generate an audio-video sequence that is actually presented to the end users.

BIFS describes an MPEG4 scene with a hierarchical structure, which can be represented with a tree (see Figure 2). Each node of the tree is an audio/video objects. In addition, the scene can be dynamic, that is, BIFS supports streaming delivery, so that the scene description can be updated dynamically. For instance, without changing the bitstream already delivered for the different objects in the above scene – the teacher, the globe, the desk, the white board, and the presentation web pages, we could change the spatial locations of the different objects, creating an animated classroom, with the direction and volume of the speech from the teacher changing with each update of the scene description. This feature allows the creation of a very complicated and lively presentation with little overhead.

In addition, an MPEG4 player will also support user interactivity at the object level. The user can click on any of the object, or move mouse inside the perimeter of an object. This event will trigger an event signal sent back to the server/encoder side. The property or composition of the scene can be changed according to this event. This object-based interactivity is very valuable for multimedia learning applications, where users can explore on their own terms and the scene adapts according to the preference of the user.

Streaming and Scalable Media Capability

Multimedia delivery over the Internet is one of the focused applications for MPEG4. There are several features that are particularly important for multimedia on the Internet. (1) The media should be able to be streamed to the end user, which means that the receiver can render the content as it receives the data without waiting for the downloading of the whole file. The flurry of deals and mergers by the Internet streaming video companies demonstrates the importance of the streaming media. (Yahoo's acquisition of Broadcast.com for $5.7 billion is an example). Streaming will also enable the delivery of *live* events without delay. (2) The media should be scalable. There are a variety of bandwidths available for different users accessing the same content. The servers and users should have the flexibility to tradeoff quality with downloading time and bandwidth requirement.

MPEG4 provides tools for addressing both of the above features. The scalability feature is supported in the still texture and video tools (scalability is a serious concern mostly only for video and image due to their large data amount). For video presentation, from the same bitstream, the user can extract a sub-bitstream that represents either a lower frame rate, or a lower resolution. This allows the users with lower bandwidth connections to trade-off either motion smoothness or video resolution for a lower bandwidth requirement. The still images coded with the MPEG4 still texture tools support both resolution and quality scalability with very fine granularity. The combination of these tools makes MPEG4 suitable for scalable media delivery over the Internet. In addition, The MPEG4 file format is designed so that the media streams from different objects can be delivered in a streaming fashion. It also supports random access, fast editing and searching of the media stream.

The JPEG2000 Image Coding Standard

The original JPEG image coding standard developed by the ISO/IEC Joint Picture Expert Group (officially named ISO-IEC/JTC1/SC29/Working Group (WG) 01) has been a big success in the market place. The JPEG image format was widely used for the Internet, digital photography, and multimedia desktop publishing along with numerous other applications. After the completion of DCT-based JPEG, many advances have been made in image coding such as wavelet and subband based technology, and many limitations and shortcomings of the JPEG standard were also discovered and widely recognized, such as severe blocking effects at low bitrate, insufficient support for scalability (both spatial and quality scalability), lack of object based and region based representation, lack of support for error resilience, etc. Recognizing these problems of the current JPEG standard and the promises of the new technologies, the ISO-IEC/JTC1/SC29/WG01 committee decided that the timing for the development of a new image coding standard was ripe. The new work item for developing the next generation image coding standard was approved by ISO/IEC SC29 subcommittee in late 1996, and a call for proposal was issued in March 1997.

The charter for the JPEG2000 standardization efforts is to develop an image coding algorithm that will meet the needs of its major application areas into the next millennium. High on the list are *superior image quality at low bitrate*, *flexibility* of the bitstream for supporting *scalability, progressive transmission, robustness to channel errors, lossless coding, and region-of-interest coding (ROI)*, preferably all achieved with similar complexity as that of the current JPEG standard.

There are totally 24 full proposals and 7 partial submissions received by the JPEG committee, with most of the proposals based on wavelet methods, and a thorough subjective and objective testing was conducted at the JPEG Sydney meeting in November 1997. A baseline codec was formed (the Verification Model), which was the anchor point to which subsequent proposals were compared. Currently, intensive work is still on going in evaluating various proposals for error resilience, entropy coding, low-complexity transform, shape coding, and region-of-interest coding. According to the current schedule, the JPEG2000 standard will become Committee Draft (CD) in 1999, at which point the technical content of the standard will

a woman was decomposed into edges and textures at different scales, and most of the coefficients after the transform become also zero (the white space).

After the wavelet transform, entropy coding is used to compress the amount of data needed to represent the image. Currently, JPEG2000 uses a block based arithmetic coder with a context switching scheme[9]. The entropy coding encodes the value of the wavelet coefficients in a bitplane-by-bitplane fashion, which ensures the feature of fine granularity scalability. Other than scalability, the coder also provides excellent support for random access, region of interest coding, error
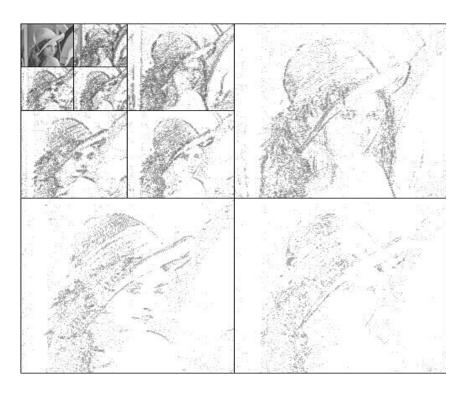

Figure 3. Wavelet Decomposition of A Picture

be stable. It will become Draft International standard (DIS) in 2000 and final International Standard (IS) in 2001. As it stands at this moment, the JPEG2000 standard is still subject to drastic changes given that almost all aspects of the current baseline algorithm are being evaluated and compared with competing proposals. However, the consensus seems to have emerged that the new JPEG standard will be based on wavelet technology (as opposed to Discrete Cosine Transform –DCT based technology in the current JPEG standard).

Wavelet transform is a mathematical tool for decomposing a signal (such as a picture) into a multi-resolution representation. It has many elegant mathematical properties so that after the transformation, the picture is much more compact in energy and much easier to compress. Figure 3 shows a wavelet transform of a picture. We can see that the original picture of

resilience, and information embedding. It will also support streaming of the image data over networks. The readers are encouraged to stay tune and a lot of information can be found at the web site: http://www.jpeg.org.

Summary

Both MPEG4 and JPEG2000 promise to produce the next generation multimedia representation and delivery standards that will significantly enhance features and functionality. The application software and equipment that use these enabling technology is expected to come to the market in the next couple of years. They promise to bring much enhanced user experience to the Internet, to CD-ROM titles, and other interactive learning venues. These new application software and web sites will, in no doubt in my mind, significantly elevate the

importance of the role multimedia plays in the informing science areas.

To probe further, the readers are referred to the following web sites:

- The official MPEG web site:
  http://www.cselt.it/mpeg
- MPEG4 system:
  http://garuda.imag.fr/MPEG4
- MPEG4 video:
  www.hhi.de/mpeg-video
- MPEG4 Synthetic coding:
  www.es.com/mpeg4-snhc
- MPEG4 audio coding:
  www.tnt.uni-hannover.de/project/mpeg/audio
- JPEG2000:
  www.jpeg.org

References

1. ISO/IEC 11172-2, "*Information Technology-Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbits/sec: Part 2: Video*", August 1993.

2. ISO/IEC 13818-2, "*Information Technology – Generic Coding of Moving Pictures and Associated Audio Information: Video*", 1995.

3. R. Talluri, "*MPEG-4 Status and Directions*", in Proceedings of the SPIE Critical Reviews of Standards and Common Interfaces for Video Information System, pp. 252—262, Philadelphia, October, 1997.

4. F. Pereira, "*MPEG-4: A New Challenge for the Representation of Audio-Visual Information*", in Proceedings of the Picture Coding Symposium, PCS'96, pp. 7-16, March 1996, Melbourne, Australia.

5. T.Sikora, "*The MPEG-4 Video Standard Verification Model", IEEE Trans. CSVT*, Vol.7, No.1, Feb.1997.

6. Rob Koenen, "*MPEG4 – Multimedia for Our Time*", IEEE Spectrum Magazine, February 1999.

7. T.Sikora and L.Chiariglione, "*The MPEG-4 Video Standard and its Potential for Future Multimedia Applications* ", Proc. IEEE ISCAS Conference, Hong Kong, June 1997.

8. ISO/IEC 14496-2, "Information Technology – Coding of Audio-Visual Objects: visual", March 1998.

9. The ISO/IEC JPEG Committee, "*JPEG2000 Verification Model 3.0 (B)*", ISO/IEC JTC1/SC29/WG11 JPEG WG1N 1137, Los Angels, November 1998.

♧