Enhanced Broadcast Services with Complimentary Delivery Over IP

Jan van der Meer
Philips Consumer Electronics

Abstract

The increase of available bandwidth over IP, the decrease of costs of storage devices and the availability of new technology for coding and streaming of audio and video is creating more and more opportunities for new enhanced broadcast services. The paper discusses technologies that enable those services and examples of applications.

INTRODUCTION

One of the technologies that is expected to play an important role in enhanced broadcast services with complimentary delivery over IP is MPEG-4, an object oriented multimedia standard that offers a variety of tools. In addition to providing tools for coding of individual video, audio and graphical objects, MPEG-4 provides also facilities to compose an MPEG-4 scene from a set of such objects. The composition may vary in time and space, which is particularly powerful in the case of multimedia applications with a behavior that is dynamic and yet unknown at application launch, for example covering a life sports event.

To enhance an MPEG-2 based broadcast service, the MPEG-4 objects can be transported over the same MPEG-2 Transport Stream as the MPEG-2 audiovisual broadcast content. Typically, this will be done for objects that need to be transported to all viewers of the broadcast service. However, transport over IP is possible too, which may in particular be useful when transport is needed to one or more individual users only, for example in case of a personalized service.

The MPEG-4 content can be played back immediately, but it is also possible to store one or more of the objects on a storage device in the Set Top Box for playback at a later time, for example upon a command of the user. In summary, delivery and playback options can be exploited as deemed appropriate by the application.

This paper first provides an introduction on important MPEG-4 features for enhancement of MPEG-2 programs with MPEG-4 content. Next, it is discussed how to incorporate MPEG-4 features in MPEG-2 broadcast services. Finally, some example applications are presented.

MPEG-4 FEATURES

MPEG-4 is an ISO/IEC standard, defined by MPEG, the Committee that earlier developed the Emmy Award winning standards known as MPEG-1 and MPEG-2, which enabled a large variety of digital video applications for Compact Discs, Broadcast, and DVD. MPEG-4, their most recent achievement, is an object based multimedia standard, offering scalability and flexibility, in combination with a high coding efficiency over a large range of bandwidth.

Though there is some functional overlap between MPEG-4 and MPEG-2, MPEG-4 will not replace MPEG-2. At the bitrates and resolutions commonly utilized by MPEG-2 applications there is no improvement in coding efficiency and hence no justification to replace MPEG-2 by MPEG-4 tools. Instead, the MPEG-4 specification is designed to offer a set of new features that can be exploited to add value to existing MPEG-2 applications. In particular the object oriented approach of MPEG-4 enables the design of sophisticated multimedia applications. An overview of MPEG-4 features for video, graphics, audio and composition of scenes is provided below.

Video

The video tools are capable of coding natural textures, images and video. In addition, arbitrary shaped objects and transparency of video objects are supported. MPEG-4 video offers scalability and flexibility in combination with a high coding efficiency over a large range of bandwidth, from as low as 5kbit/s up to about 1 Mbit/s. This makes MPEG-4 especially suitable for video streaming applications in environments where the available bandwidth may vary.

Coding of natural images and video is achieved in a way similar to conventional MPEG-1 and MPEG-2 coding, using motion compensated predictive DCT technology, with a high level of flexibility with respect to input formats, frame rates, pixel depth and bitrates. When compared with MPEG-1 and MPEG-2, motion compensation technologies are enhanced to support very low bitrates. As a result, MPEG-4 is extremely efficient at very low bitrates, making the MPEG-4 standard the obvious choice for streaming media over the Internet.

In addition to the MPEG-1 and MPEG-2 capabilities, MPEG-4 supports arbitrary shaped objects and transparency of video objects. MPEG-4 also includes a dedicated coding scheme for textures and still images based on a very efficient and scalable zerotree wavelet algorithm. Furthermore the MPEG-4 standard supports some 2D and 3D modeling techniques, as well as synthetic objects, in particular human face and body animation.

Graphics

MPEG-4 defines a rich set of 2D and 3D graphical functions, largely based on VRML. For applications requiring low-complexity graphics, a 2D Graphics profile is defined. For transport of graphical data an efficient binary format has been specified that compresses VRML 2.0 data with a factor of typically 8 - 15. This format also allows for very low bitrate animations, typically at the bitrate of a few kbit/second.

Audio

The MPEG-4 audio tools are capable of coding speech and music over a wide range of bit rates and sampling frequencies. Three different types of codecs are defined. The lowest bitrate range is covered by parametric coding techniques; 2 - 4 kbit/s for speech with a 8 kHz sampling frequency and 4 - 16 kbit/s for music with a sampling frequency of 8 or 16 kHz. Speech coding at the medium bitrates between about 6 - 24 kbit/s uses Code Excited Linear Predictive (CELP) coding techniques. In this region, two sampling rates, 8 and 16 kHz, are used to support both narrowband and wideband speech, respectively. For bitrates starting below 16 kbit/s, and typically up to 128 kbit/s for stereo, Time to Frequency (T/F) coding techniques are applied, with sampling frequencies such as 8, 16, 24, 32 and 48 kHz.

To support audio streaming over Internet with a range of available bandwidth and a variable Quality of Service (QoS), MPEG-4 defined the Scalable Audio Profile. Using this profile, it is possible to increase the audio quality seamlessly when more bandwidth comes available and to decrease the audio quality gracefully when the available bandwidth decreases.

Scenes

MPEG-4 is an object oriented multimedia standard. In addition to providing support for the coding of individual video, audio and graphical objects, MPEG-4 also provides facilities to compose an MPEG-4 scene from a set of such objects. The composition may vary in time and space, which is particularly powerful in the case of multimedia applications with a dynamic behavior. Examples thereof are applications to enhance the broadcast of life sport events. During such life broadcasts, several situations may happen that are worthwhile to be reflected in an immediate change of the scene composition of the enhanced broadcast application. MPEG-4 allows for example to add and remove on the fly video objects simultaneously overlayed on the screen.

The necessary composition information forms the scene description that is constructed using the so-called BIFS (Binary Format for Scene description). Composition information is coded and transmitted together with the media objects. BIFS provides MPEG-4 with a rich set of scene construction operators, including the VRML graphics primitives that can be used to construct sophisticated scenes.

Delivery of MPEG-4 content over MPEG-2 Systems and over IP

MPEG-4 is an abstract standard that does not define transport mechanisms; for interoperable services additional specifications are needed to define transport of MPEG-4 data. The MPEG Committee defined how to carry MPEG-4 content over MPEG-2 System streams in Amendment 7 to the MPEG-2 System specification, while the IETF defines carriage of MPEG-4 over IP in a joint effort with the MPEG Committee.

ENHANCED MPEG-2 PROGRAMS WITH MPEG-4 CONTENT

Within MPEG-2 Transport Streams, the Program Map Table, PMT, is used to define which elementary streams form a program. Within the PMT also reference can be made to MPEG-4 content. The MPEG-4 content may represent an individual MPEG-4 elementary stream, or an MPEG-4 scene with one or more MPEG-4 objects.

Individual MPEG-4 elementary streams

An individual MPEG-4 stream may represent for example an MPEG-4 encoded complementary speech channel or a low frame rate video of a small size for overlay on full screen MPEG-2 video. Each individual MPEG-4 audio and visual stream is carried in PES packets. In the PES header, PTSs are encoded in the same way as for MPEG-2 elementary streams, based on the MPEG-2 System Time Clock. In this way the decoding and presentation of the individual MPEG-4 elementary stream is defined directly in terms of the MPEG-2 time base.

MPEG-4 scenes and objects

To identify the MPEG-4 objects that are to be composed into an MPEG-4 scene, a unique ID, the ES_ID, is assigned to each such MPEG-4 object. An MPEG-4 object may represent audio, video, text, graphics, or other content. An MPEG-4 object is not required to be MPEG-4 encoded. For example, also an MPEG-2 video or audio stream can be an MPEG-4 object.

MPEG-4 System streams

MPEG-4 audio and visual elementary streams can be carried directly in PES as individual MPEG-4 elementary stream. Next to this method, also MPEG-4 System tools can be used to carry MPEG-4 content over MPEG-2, in particular SL-packetized streams and FlexMux streams.

In MPEG-4 Systems, SL packets are the basic entity for carriage of access units. Each SL packet carries exactly one access unit or a part thereof. The header of the SL packet contains time stamps and other data for the contained access unit. A sequence of SL packets with data from the same elementary stream is called a SL-packetized stream.

The MPEG-4 FlexMux tool is capable of multiplexing SL-packetized streams into a FlexMux stream. A FlexMux stream consists of a sequence of FlexMux packets. Each SL-packetized stream in a FlexMux has its own FlexMux channel, identified by the FlexMux channel number coded in the header of the FlexMux packet.

MPEG-4 time base(s)

In principle, each MPEG-4 object has its own time base. Elementary streams carried in PES without the use of MPEG-4 System streams are locked to the MPEG-2 System Time Clock, STC, in the same way as any MPEG-2 audio or video stream that is part of the same program. MPEG-4 elementary streams that are carried using MPEG-4 System streams have a time base with the following characteristics :

- The object time base is locked to the MPEG-2 STC;
- There is a fixed time offset between the object time base and the MPEG-2 STC.

In this case the object time base is either carried by the SL-packet header or by a specific FlexMux channel. The time offset between the object time base is defined through the use of MPEG-2 and MPEG-4 time stamps. See amendment 7 to the MPEG-2 System specification.

Complementary delivery over IP

Once developed, an MPEG-4 application requires transport of the MPEG-4 content. If the streams are intended for broadcast to many clients, transport over MPEG-2 may be most suitable, but content that is delivered to a single client may be delivered more efficiently over IP. In any case both delivery methods are available as complementary options which can be exploited as appropriate.

Delivery of SL-packetized streams and FlexMux streams over IP is specified by IETF, in a joint effort with the MPEG Committee. MPEG-4 allows the content to be authored independently of the delivery method. The application requires that the content has been authored under the same conditions as for delivery over MPEG-2, and therefore the format of constructed SL-packetized and FlexMux streams is fully transparent to transport over MPEG-2 or over IP. However, applications should take into account that the end-to-end delay for delivery over IP may be significantly larger.

EXAMPLES OF APPLICATIONS

MPEG-4 Scene composition is a powerful tool to design multimedia applications. It specifies how video, audio and graphics and other objects relate in time and space. In this section four examples are given of multimedia applications designed to enhance digital broadcast services.

Audience attractor

Assume a Pay TV broadcast of a tennis event. Instead of the usual way of reporting from a tennis match with a broadcast of a single full screen video stream and a single audio stream, the broadcast is now composed of multiple objects :

- two tennis players as two foreground video objects;
- the tennis stadium as a background video object;
- an audio object with a commentary voice;
- the ambient sound in the stadium as a background audio object.

The PayTV coverage of the event is broadcasting the background objects, the tennis stadium with its ambient sound for free, to attract the target audience. Upon payment, the tennis players and the commentary voice are added to the background.

Personalized advertizing

Assume further to the previous example some advertisement boards as texture objects within the background object The advertisements on the boards depend on the geographic location and profile of the consumer, to reach the target audience for the advertisements.

Immersive sport coverage

Assume a major sport event such as the Olympics, World Football Championship, Tour de France Cycling or Car Racing. The event is covered with one or more digital broadcasts. Each broadcast consists of the usual full screen video and associated audio streams, but in addition other streams are provided too, such as one or more of the following small size pictures that can be overlayed on the full screen video. These overlayed pictures may or may not have associated audio, and may represent for example one or more of the following :

- Coverage of a simultaneously ongoing game
- Status of the race at different positions
- Performance of a specific racer
- Highlights of the game or race until now
- Summary of a previous race or game
- Replay of something that happened during the game or race.

The additional content may be available in many ways. It may be broadcast or stored locally or at a server. In case the additional content is intended for a broad audience, it is likely delivered over the broadcast channel, but in case of delivery to a single user or a small group of users, the content may be delivered more efficiently over IP.

The objective is to optionally offer users an enhanced coverage of the event. If the user wishes, after tuning in he can ask for the highlights until now, he can keep an eye on the progress of simultaneous games, have a look to the performance of his favorite racer and have a replay of what happened earlier. The user may be able to choose between watching the event in the usual lean back position or to have an exciting multimedia experience in a little bit more lean forward position or anything in between, as desired.

Enhanced EPG

Assume a bouquet of digital broadcast services offered over a medium such as cable. For user convenience and promotional purposes an enhanced EPG is offered; the EPG can be overlayed on top of any of the selected broadcast programs, exploiting the transparency and arbitrary shape features of MPEG-4. The EPG shows a vertical axis that lists all available services and a horizontal axis that represents time. The user can very conveniently select a service and a point in time and will get presented in a small window information on what is broadcast by the selected service at the selected time. The information can be broadcast, stored locally or provided over IP. The user can select programs of a certain type and when available, such programs are recommended to the user, either to watch or to store for later use.

CONCLUSION

The MPEG-4 standard provides a truly open specification with a large variety of features enabling powerful multimedia applications. MPEG-4 provides solutions for delivery over Internet with its largely varying delivery rates, but MPEG-4 also provides a solution for high QoS services such as provided by digital broadcasts. Using the same MPEG-4 standard, applications can exploit the strengths of both delivery methods. Complete new ways of broadcast become possible.

Jan van der Meer
Technology Manager
Philips Consumer Electronics
Email : jan.vandermeer@philips.com