CS 1st yr(Multimedia & Animation)

Unit 3

MPEG-4

The MPEG-4 object-based representation approach goes beyond the limits of previous MPEG product standards by opening new frontiers in the way users will play with, create, reuse, access, and consume audiovisual content. Learn how MPEG-4 and its associated technology provide the means to launch a great diversity of applications for the emerging third-generation mobile networks, for the Internet, and even for digital radio and cable broadcasting networks.

Introduction

Keywords: audiovisual object-based representation, audiovisual objects, MPEG-4, standardization process

"What does it mean, to see? The plain man's answer (and Aristotle's, too) would be, to know what is where by looking. In other words, vision is the process of discovering from images what is present in the world, and where it is."

The first video coding standards and their underlying representation models mainly address the vision process by providing video representation in the form of a sequence of rectangular 2D frames, giving users a window to the real world: the television paradigm. However, the process of vision is often just the initial part of the task at hand, because typically humans need and want to see, to take actions after, to interact with the objects identified. A similar reasoning can be made regarding the process of hearing and the corresponding audio representation models.

Although the television paradigm dominated audiovisual communications for many years, the situation has been evolving quickly in terms of the ways audiovisual content is produced, delivered, and consumed. Moreover, hardware and software are becoming increasingly powerful, with microelectronic technology providing new programmable processors, opening new frontiers to the representation technologies used and to the functionalities provided.

Producing content today is easier than ever before. Digital still cameras directly storing in JPEG format have hit the mass market. Together with the first digital video cameras directly recording in MPEG-1 format, this represents a major step for consumer acceptance of digital audiovisual acquisition and compression technology. This step transforms every individual into a potential content producer, capable of creating content that can be easily distributed and published on the Internet. In addition, more content is being synthetically produced—that is, computer generated—and integrated with natural material in truly hybrid audiovisual content. The various

pieces of content, digitally encoded, can be successively reused without the quality losses typical of the previous analog processes.

Whereas audiovisual information, notably the visual portion, until recently was carried only over very few networks, the trend is now toward the generalization of visual information in every single network. Moreover, the increasing mobility in telecommunications is a major trend. Mobile connections will not be limited to voice; other types of data, including real-time media, are already emerging. Because mobile telephones are replaced every two to three years, new mobile devices can finally make the decade-long promise of audiovisual communications a reality. Notice that the need for visual communication is much stronger when people are not at home, and so have something to show besides the usual living room—for example, the nice beach where they are vacationing. This reinforces the relevance of audiovisual mobile communications.

The explosion of the World Wide Web and the acceptance of its interactive mode of operation have clearly shown that the traditional television paradigm will no longer suffice for audiovisual services. Users want to have access to audio and video as they now have access to text and graphics. This requires moving pictures and audio of acceptable quality at low bit rates on the Web, and Web-type interactivity with live content. It will be possible to activate relationships between users (in a potentially virtual world) through hyper linking—the Web paradigm—and to experience interactive immersion in natural and virtual environments—the games paradigm.

As many of the emerging audiovisual applications demanded interworking, the need to develop an open and timely international standard addressing the needs mentioned above became evident. In 1993, MPEG (Moving Picture Experts Group) [MPEG] launched the MPEG-4 project, later formally called *Coding of Audio-Visual Objects*, to address (among other things) the requirements associated with the new applications resulting from these trends.

The need for any standard comes from an essential requirement relevant for all applications involving communication between two or more parts: *interoperability*. Interoperability is thus the requirement expressing the user's dream of exchanging any type of information without any technical barriers, in the simplest way. Without a standard way to perform some of the operations involved in the communication process and to structure the data exchanged, easy interoperability between terminals would be impossible. Having said that, it is clear that a standard shall specify the minimum number of tools needed to guarantee interoperability (because it is important that as many as possible non-normative technical zones exist), to allow the incorporation of technical advances, and thus to increase the lifetime of the standard, as well

as to stimulate industrial technical competition. The existence of a standard also has important economic implications, because it allows the sharing of costs and investments and the acceleration of applications deployment.

MPEG has been responsible for the successful MPEG-1 (ISO/IEC 11172) and MPEG-2 (ISO/IEC 13818) standards, which have given rise to widely adopted commercial products and services, such as Video-CD, DVD, digital television, digital audio broadcasting (DAB), and MP3 (MPEG-1 Audio layer 3) players and recorders. The MPEG-4 standard (ISO/IEC 14496) is aimed at defining an audiovisual coding standard to address the emerging needs of the communication, interactive, and broadcasting service models as well as the needs of the mixed service models resulting from their convergence. The apparent convergence of the three traditionally separate application areas—communications, computing, and TV/film/entertainment—was evident in their cross-fertilization, with functionalities characteristic of each area increasingly emerging in the others (e.g., personal communications including video information or entertainment including interactive capabilities).

Following the previous successes—in fact, as a natural consequence of the vision underpinning MPEG-4—MPEG initiated in 1996 another standardization project addressing the problem of describing audiovisual content to allow the quick and efficient searching, processing, and filtering of various types of multimedia material: MPEG-7 (ISO/IEC 15938), officially called *Multimedia Content Description Interface* [N4509]. In fact, digital audiovisual information is more and more accessible to everyone, not only in terms of consumption but also in terms of production. But if it is much easier today to acquire, process, and distribute audiovisual content, it must be equally easy to access the available information, because huge amounts of audiovisual information are being generated all over the world every day. The need for a powerful way to quickly and efficiently identify, search, and filter various types of audiovisual content, by humans or machines (also using non-text-based technologies), directly follows from the urge to efficiently use the available audiovisual content and the difficulty of doing so. MPEG-7 will specify a standard way of describing various types of audiovisual information, including still pictures, video, speech, audio, graphics, 3D models, and synthetic audio, regardless of their representation format (e.g., analog or digital) and storage support (e.g., paper, film, or tape). In comparison with other available or emerging solutions for audiovisual content description, MPEG-7 can be mainly distinguished by (a) being general purpose, meaning its ability to describe content from many application environments; (b) its object-based representation model, meaning the capability of independently describing individual objects within a scene (be it MPEG-4 or any other format); (c) the integration of low-level and high-

level features/descriptors into a single description framework, allowing it to combine the power of both types of descriptors; and (d) its extensibility, provided by the Description Definition Language, which allows MPEG-7 to keep growing, to be extended to new application areas, to answer newly emerging needs, and to integrate novel description tools [PeKo99]. The MPEG-7 standard was finalized in the summer of 2001.

Following the development of standards addressing more focused targets, MPEG acknowledged the lack of a big picture that described how the various elements building the infrastructure for the deployment of applications using multimedia content relate to each other, or even if there are missing standard specifications for some of these elements. To address this problem, MPEG started the MPEG-21 project (first ISO/IEC 18034, now ISO/IEC 21000), formally called *Multimedia Framework*, with the aim of understanding if and how these various elements fit together, and to discuss which new standards might be required if gaps in the infrastructure exist. Once this work has been carried out, new standards will be developed for the missing elements with the involvement of other bodies, where appropriate; finally, the existing and novel standards will be integrated in the MPEG-21 multimedia framework. The MPEG-21 vision is thus to define a multimedia framework to enable transparent and augmented use of multimedia resources across a wide range of networks and devices used by different communities. The MPEG-21 multimedia framework will identify and define the key elements needed to support the multimedia value and delivery chain, as well as the relationships between and the operations supported by them.

After briefly covering the context that motivated the birth of the MPEG-4 project, this chapter presents its major objectives in terms of functionalities, requirements, tools, and applications, as well as its organization and the sequence followed to achieve the goals defined. This chapter also addresses the MPEG modus operandi, notably its mission, principles, and specific approach to the development of standards. Finally, the objectives and working approach of the MPEG-4 Industry Forum will be presented.

1) MPEG-4 Objectives

Although MPEG discussions about projects beyond MPEG-2 began as early as May 1991, at the Paris MPEG meeting, it was not until September 1993 that the MPEG Applications and Operational Environments (AOE) subgroup was set up and met for the first time. The main task of this subgroup was to identify the applications and requirements relevant to the far-term, very low bit-rate coding solution to be developed by International Organization for Standardization (ISO)/MPEG as stated in the initial MPEG-4 project description [N271]. At the same time, the near-term hybrid coding solution being developed within the International Telecommunications

Union-Telecommunications Standardization Sector (ITU-T) Low Bit-rate Coding (LBC) group started producing the first results (later, the ITU-T H.263 standard [H263]). It was then generally felt that those results were close to the best performance that could be obtained by block-based, hybrid, DCT/motion-compensation video coding schemes.

In July 1994, the MPEG meeting marked a major change in the direction of MPEG-4. Until that meeting, the main goal of MPEG-4 was to obtain a significantly better compression ratio than could be achieved by conventional coding techniques. Few people, however, believed it was possible, in the next five years, to make enough improvements over the LBC standard (H.263 and H.263+) to justify a new standard.[2] So the AOE subgroup was faced with the need to broaden the objectives of MPEG-4, believing that pure compression gains would not be enough to start a new MPEG standardization project. The subgroup then began an in-depth analysis of the audiovisual world trends, based on the convergence of the TV/film/entertainment, computing, and telecommunications worlds. The conclusion was that the emerging MPEG-4 coding standard should support new ways (notably content-based) of communicating, accessing, and manipulating digital audiovisual data.

2) Functionalities

Following this change of direction and the analysis made, the vision driving the MPEG-4 standard was explained through the eight new or improved functionalities described in the MPEG-4 Proposal Package Description (PPD), prepared by the time of the first MPEG-4 call for proposals in July 1995. These eight functionalities came from an assessment of the functionalities that would be useful in future audiovisual applications, but which were not supported (or at least not well supported) by the available coding standards. The eight new or improved functionalities were clustered in three classes related to the three worlds—TV/film/entertainment, computing, and telecommunications—the convergence of which MPEG-4 wanted to address:

1. **Content-based interactivity**

    - **Content-based multimedia data access tools:** MPEG-4 shall provide efficient data access and organization based on the audiovisual content. Access tools may be indexing, hyper linking, querying, browsing, uploading, downloading, and deleting. Sample uses include content-based retrieval of information from online libraries and travel information databases.

- **Content-based manipulation and bit stream editing:** MPEG-4 shall provide syntax and coding schemes to support content-based manipulation and bit stream editing without the need for transposing. This means the user should be able to select one specific object in the scene/bit stream and change some of its characteristics. Sample uses include home movie production and editing, interactive home shopping, and the insertion of a sign language interpreter or subtitles.

- **Hybrid natural and synthetic data coding:** MPEG-4 shall support efficient methods for combining synthetic scenes with natural scenes (e.g., text and graphics overlays), the ability to code and manipulate natural and synthetic audio and visual data, and decoder-controllable methods of mixing synthetic data with ordinary video and audio (allowing for interactivity). For example, in virtual reality applications, animations and synthetic audio (e.g., MIDI) can be mixed with ordinary audio and video in games, and graphics can be rendered from different viewpoints.

- **Improved temporal random access:** MPEG-4 shall provide efficient methods to randomly access, within a limited time and with fine resolution, parts (e.g., frames or objects) from an audiovisual sequence. Example usage: audiovisual data can be randomly accessed from a remote terminal over limited-capacity media, a fast-forward can be performed on a single audiovisual object in the sequence.

2. **Compression efficiency**

- **Improved coding efficiency:** The growth of mobile networks creates an ongoing demand for improved coding efficiency; therefore, MPEG-4 set as its target providing subjectively better audiovisual quality compared to existing or other emerging standards (such as H.263), at comparable bit rates. Sample uses include efficient transmission of audiovisual data on low-bandwidth channels and efficient storage of audiovisual data on limited-capacity media, such as chip cards.

- **Coding of multiple concurrent data streams:** MPEG-4 shall provide the ability to efficiently code multiple views/soundtracks of a scene as well as sufficient synchronization between the resulting elementary streams.[4] For stereoscopic and multiview video applications, MPEG-4 shall include the ability to exploit redundancy in multiple views of the same scene, also permitting solutions that allow compatibility with normal video. Sample uses include multimedia entertainment (e.g., virtual reality

games and 3D movies), training and flight simulations, multimedia presentations, and education.

3. **Universal access**

- **Robustness in error-prone environments:** Because universal accessibility implies access to applications over many wireless and wired networks and storage media, MPEG-4 shall provide an error robustness capability. Particularly, sufficient error robustness shall be provided for low bit-rate applications under severe error conditions. Sample uses include transmission from a database over a wireless network, communicating with a mobile terminal, and gathering audiovisual data from a remote location.

- **Content-based scalability:** MPEG-4 shall provide the ability to achieve scalability with a fine granularity in content, spatial resolution, temporal resolution, quality, and complexity. Content-scalability may imply the existence of a prioritization of the objects in the scene. Sample uses include user selection of decoded quality of individual objects in the scene and database browsing at different scales, resolutions, and qualities.

These functionalities were essential to shaping the MPEG-4 vision, balancing completely new functionalities with more traditional ones, and thus allowing a bridge from the past to the future not only in terms of functionalities but also in terms of tools and experts.

3) Requirements

Following the identification of the fundamental MPEG-4 functionalities, MPEG started a requirements development process, which has been continuously evolving since then. The requirements serve to drive the tools development process and assure that the right technology is being specified: Tools will be developed to fulfill the identified requirements, and no tools that do not address any requirement will be defined.

By the middle of 2001, the MPEG-4 requirements were structured as shown in Table. The requirements are organized in terms of major technical areas, which do not directly correspond either to Parts of the MPEG-4 standard or to MPEG working subgroups. Within each category, the requirements are to be fulfilled by the set of tools selected for the standard, and not all requirements must be addressed by each individual tool. It is the right combination of tools that allows building the algorithms that can address the specific needs of a certain class of applications.

MPEG-4 requirements

Requirements for systems	
Flexibility Multiplexing of audio, visual, and other information Composition of audio and visual objects Application texture Downloading User interaction Media interworking Compatibility Robustness to information errors and losses Object-based bitstream manipulation and editing	Multipoint operation Object content information (OCI) Video-related metadata Delay Configuration modes Priority of audiovisual objects Dynamic resource management Reference to associated MPEG-7 data File format Textual format
Requirements for natural video objects	

Object-based representation Video content Object-based bitstream manipulation and editing Object-based random access Object quality and fidelity Coding of multiple concurrent data streams Robustness to information errors and losses	Object-based coding flexibility Object-based scalability Delay modes Formats Bit-rate modes Complexity modes Still images Tandem coding[5]
Requirements for synthetic video objects	
Types of synthetic video objects 2D/3D mesh compression Definition and animation parameter	Text overlay Image and graphics overlay View-dependent texture scalability Geometrical

compression Texture mapping	transformations Video object tracking
Requirements for natural audio objects	
Object-based representation Audio content Object-based bitstream editing and manipulation Object-based scalability Object-based random access and user controls Time scale change Pitch change	Robustness to information errors and losses Delay modes Complexity modes Bit-rate modes Downmix[6] Transcoding Tandem coding Audio formats Improved coding efficiency
Requirements for synthetic audio objects	
Low bit-rate speech Synthetic speech data	Text to speech Sound synthesis

Requirements for delivery multimedia integration format (DMIF)	
Connectivity Transparency Application service enablement	End-to-end Quality of Service (QoS) management Network-based stream processing and management
Requirements for MPEG-J	
Functional requirements Byte code delivery Authentication	Byte code execution Event mechanism
Requirements for multiuser environments	
Scene graph representation Audiovisual objects and avatars representation Delivery and stream	IPMP and sharing tools Application programming interfaces

management	
Requirements for animation framework	
Enhanced texture mapping Animation support High-level shape representation	Reusability of scene graph nodes and animation streams Persistence Compression of animated objects
Requirements for intellectual property management and protection (IPMP)	
Identification of intellectual property Intellectual property management and protection hooks	Intellectual property management and protection interfaces

Although efficient compression was not the only first-priority requirement in MPEG-4 (as it had been for MPEG-1 and MPEG-2), it is clear that it has been a central requirement in MPEG-4 in the sense that, whatever the type of data that had to be represented by binary encoding (e.g., shape for video objects, facial animation parameters for 3D facial models, or even scene composition data), the target was always to reach the smallest number of bits for a certain level of quality. Although some people claim that efficient compression of data is not a must today

because of the growing availability of bandwidth, MPEG always acknowledged that bandwidth resources (either for transmission or for storage) are still limited, and thus efficient compression is required. This is even truer for some relevant recent transmission cases, such as the Internet and mobile networks, where bandwidth limitations and efficient compression are major issues. However, in the context of MPEG-4, efficient compression must be balanced against other major requirements, such as those related to interactivity capabilities (which have a price in terms of compression efficiency compared with non interactive representation schemes), if new functionalities not supported by frame-based coding schemes are to be provided.

As noted, the MPEG-4 requirements' development process has been evolving since the beginning of the standardization process; this evolution has been targeting the inclusion of additional requirements related to functionalities, which fit and complement well the MPEG-4 vision. Because the MPEG-4 standardization process is not yet finished, it cannot be said that all the requirements have been addressed by means of an MPEG-4 tool. However, it is possible to say that all of the not-yet-addressed requirements are either being worked on or should be removed in a short time if the industries do not show sufficient support to move to the technical development phase.

1.1.3 Tools

The major trends mentioned—notably the mounting presence of audiovisual media on all networks, increasing mobility, and growing interactivity—have driven, and continue to drive, the development of the MPEG-4 standard.

To address the identified functionalities and requirements [N4319], a set of tools was developed to perform the following functions [Pere99]:

- Efficiently represent a number of data types through media codecs
    - Video from very low bit rates to very high-quality conditions
    - Music and speech data for a very wide bit-rate range, from transparent music to very low bit-rate speech
    - Generic dynamic 3D objects as well as specific objects such as human faces and bodies
    - Synthetic speech and music, including support for 3D audio spaces
    - Text and graphics

- Provide fine granularity scalability in the quality, temporal, and spatial dimensions
- Provide, in the encoding layer, resilience to residual errors for the various data types, especially under difficult channel conditions such as mobile ones
- Independently represent the various objects in the scene, allowing independent access for their manipulation and reuse
- Compose audio and visual (natural and synthetic) objects into one audiovisual scene in a synchronized way
- Describe the objects and the events in the scene
- Provide interaction and hyper linking capabilities
- Manage and protect intellectual property on audiovisual content and algorithms, so that only authorized users have access to the content
- Provide a delivery-media-independent representation format, to transparently cross the borders of different delivery environments

The major difference with previous audiovisual coding standards, at the basis of the new functionalities, is the object-based audiovisual representation model that underpins MPEG-4. An object-based scene is built

MPEG-4 object-based representation architecture

now feasible. There are many more advantages—such as the selective spending of bits, the easy reuse of content without transposing, and the provision of sophisticated coding solutions for scalable content on the Internet—all of them resulting from the adoption of the object-based representation model.

The applications that benefit from the new concepts and functionalities are found in many (and very different) environments. Therefore, MPEG-4 is constructed as a toolbox rather than a monolithic standard, using profiles that provide solutions for these different application settings. This means that, although MPEG-4 is a standard comprising a vast array of technologies, it is structured so that solutions are available at the measure of the needs. It is the task of each implementer to extract from the MPEG-4 standard the technical solutions adequate to his or her

needs (likely a small subset of the standardized tools) by choosing the adequate profiling combination.

Because a standard is always a constraint on freedom, it is important to make it as minimally constraining as possible. To MPEG, this means a standard must offer maximum advantages by specifying the minimum necessary, allowing for competition and evolution of the technology in the *non-normative areas;* these non-normative areas correspond to the tools for which normative specification is not essential for interoperability. On the contrary, normative tools are those defined by the standard whose specification is essential for interoperability. For example, whereas video segmentation and rate control are non-normative tools, the decoding process needs to be normative. The strategy of specifying the minimum for maximum usability ensures that good use can be made of the continuous technical improvements in the relevant technical areas. The consequence is that better non-normative tools can always be used, even after the standard is finalized, and it is possible to rely on competition for obtaining ever-better results. In fact, it is through the non-normative tools that products will distinguish themselves, which only reinforces the importance of this type of tools.

1.1.4 Applications

MPEG-4 wants to address a wide range of applications, many of them completely new, as there are very new functionalities to take benefit from, and many others improved regarding those already available. Unlike MPEG-2 where the *killer application* was digital television, in a first approach just understood as the digital translation of the rather old analog version, MPEG-4 does not target a major and exclusive killer application but opens many new frontiers. Playing with audiovisual scenes and creating, reusing, accessing, and consuming audiovisual content will become easier and more powerful. New and richer applications can be developed, for example, in enhanced broadcasting, remote surveillance, personal communications, games, mobile multimedia, and virtual environments. MPEG-4 allows services combining the traditionally different service models: broadcast, (online) interaction, and communication. As such, MPEG-4 addresses *convergence,* understood as the proliferation of audiovisual information in all kinds of services and on all types of (access) networks.

The MPEG-4 Applications document describes application examples benefiting from the MPEG-4 technology that will serve as inspiration to the industry to create many more exciting applications. In this sense, MPEG-4 is a technical playground where many application constructions may be built by the manufacturers and service providers. The MPEG-4 Applications document suggests the following applications, using both audio and visual information or just one of them: broadcast, collaborative scene visualization, content-based

storage and retrieval, digital amplitude modulation (AM) broadcasting, digital television set-top box, DVD, infotainment, mobile multimedia, real-time communications, streaming video on the Internet/Intranet, studio and television postproduction, surveillance, and virtual meetings.

REETA

28/03/2020