TECHNOLOGIES FOR DIGITAL MULTIMEDIA COMMUNICATIONS: AN EVOLUTIONAL ANALYSIS OF MPEG STANDARDS

Fernando Pereira

Instituto Superior Técnico – Instituto de Telecomunicações
Av. Rovisco Pais, 1049-001 Lisboa, Portugal
fp@lx.it.pt

ABSTRACT

Multimedia communications play a growing role in the every day's life of modern societies. This change was largely made using technology specified by the MPEG standardization body. For example, the MP3 music coding format is not anymore just a technological term but a multimedia commodity that all generations are already familiar with.

This paper provides an evolutional overview of MPEG standards, discussing and explaining why certain choices were made, and thus a certain vision of the multimedia world was followed.

I. INTRODUCTION

Digital multimedia communications, from television, videotelephony, and Internet and mobile streaming to digital storage and music downloading, are nowadays a central part of modern life. It is largely recognized that MPEG standards, this means the standards developed by the ISO/IEC [1] Moving Picture Experts Group (MPEG) [1], have played and still play a major role in the starting and development of multimedia communications since they have showed the value of interoperability in the context of this type of applications. For this reason, this paper will focus on the evolutional analysis of MPEG standards, trying to explain why a certain standardization path has been followed.

When MPEG was started, in 1988, several important technologies were becoming mature enough to open new ways to deliver multimedia content to end users. Among them were audio and image/video compression, VLSI (Very Large Scale Integration) technology, optical storage, and high-speed delivery of digital information over phone lines. This fact was recognized by some consumer electronics and telecommunications companies, which had the vision that by setting standards for audio and video coding, they would create a market from which they could all benefit. They believed that providing interoperability was crucial, and that would not reduce their chances to develop successful products, but rather the opposite. Moreover this interoperability would be provided without preventing competition and excellence, by specifying only the minimum number of tools for interoperability and leaving open space for the compatible products to distinguish themselves.

Based on this vision, MPEG has set many widely used standards since its establishment. Following the evolution of multimedia applications, the group has expanded its scope from basic coding technologies to technologies supporting and complementing the audio and video compression formats such as synchronization, multiplexing, composition, graphics, metadata, and intellectual property management and protection. This process of answering to the industry needs, already started with MPEG-1, when MPEG realized that just setting audio and video compression formats would not suffice since something more was needed to support synchronization, storage and delivery. The same reasoning underpins the evolution of MPEG standards which always encompass a (growing) set of technologies targeting the provision to the user of the necessary set of tools to build increasingly more complex products for the envisioned application scenarios.

MPEG's activities have been organized around large projects, typically driven by major application domains, or functionalities. This approach led to the following set of MPEG standards:

- ISO/IEC 11172 (MPEG-1), "Coding of Moving Pictures and Associated Audio at up to about 1.5 Mbit/s", mostly addressing CD-ROM digital storage [3];
- ISO/IEC 13818 (MPEG-2), "Generic Coding of Moving Pictures and Associated Audio", mostly addressing digital television and digital storage [4][5];

- ISO/IEC 14496 (MPEG-4), “Coding of Audio-Visual Objects”, providing new object-based functionalities, synthetic and natural integration, new forms of interaction, etc [6][7];
- ISO/IEC 15938 (MPEG-7), “Multimedia Content Description Interface”, providing multimedia content description capabilities for a large range of applications [8][9];
- ISO/IEC 21000 (MPEG-21), “Multimedia Framework”, providing an integration framework for the previous MPEG standards and missing technologies such as intellectual property management and protection (IPMP), rights expression, and content adaptation [10][11].

Since the very beginning, all MPEG standards including media representation technologies (this does not happen for MPEG-21) were structured and addressed their objectives on the basis of a ‘trilogy’ of three major parts: Systems, Video/Visual and Audio. This fact, rather uncommon in other standardization bodies, shows the early recognition that multimedia is about media integration (and multimodalities) and thus considering individual media isolated cannot lead to the best multimedia standards.

More recently, in July 2005, MPEG acknowledged that the previous standardization approach based on large projects such as those mentioned above was not anymore the most adequate, but instead several smaller standards needed to be developed. Also most of these standards were not directly related to any of the large MPEG projects already in existence. Following this recognition, a new type of MPEG standards structure was created to complement the available set of large projects, notably:

- ISO/IEC 23000 (MPEG-A), “Multimedia Application Formats”, defining application driven file formats across MPEG standards [12][13].
- ISO/IEC 23001, “MPEG Systems Technologies”, defining systems tools that can be used across all MPEG large standards or, at least, are not directly linked to them.
- ISO/IEC 23002, “MPEG Visual Technologies”, defining visual coding tools that can be used across all large MPEG standards or are, at least, not directly linked to them; for example, part 1 includes a specification for the IDCT accuracy that can be referenced in substitution of the IEEE 1180 standard (which has been withdrawn by IEEE).
- ISO/IEC 23003, “MPEG Audio Technologies”, defining audio coding tools that can be used across all large MPEG standards or are, at least, not directly linked to them; for example, part 1 specifies the MPEG Surround standard, a coding standard for multichannel, spatial (typically, 5.1 channels) sound which requires the transmission of a compressed stereo (or even mono) audio program and an additional low-rate side-information channel.
- ISO/IEC 23004 (M3W), “MPEG Multimedia Middleware”, improving application portability and interoperability through the specification of a set of APIs (syntax and expected execution behavior) dedicated to multimedia as well as providing a standard way of delivering the implementation(s) of these APIs.

From these new standards, only MPEG-A will be considered in more detail in the following since it is the one at a more advanced stage. The new standards will be developed in parallel with the existing large standards which may still grow whenever this is the appropriate approach. For example, the most important MPEG video coding activities, in 2006, target the specification of the so-called Scalable Video Coding (SVC) and Multiview Video Coding (MVC) standards which will be defined as amendments (extensions) to MPEG-4 Advanced Video Coding (part 10) since the new standards are backward compatible extensions to the coding solutions already in that part of MPEG-4.

II. MPEG-1: CODING OF MOVING PICTURES AND ASSOCIATED AUDIO AT UP TO ABOUT 1.5 MBIT/S

The MPEG-1 standard [3] was developed in the period from 1988 to 1991 and represents the first generation of the MPEG family. After the ITU-T H.261 recommendation on video coding targeting videotelephony and videoconference was finalized, it became rather clear that the same video coding technology could provide the basis for a digital alternative to the broadly spread analogue video cassette player. MPEG-1 targets to bring audiovisual (AV) storage to the digital world by providing a complete audiovisual digital coding solution for digital storage media such as CD, DAT, optical drives and Winchester discs. Since CD was defined as the major target, the standard was optimized for an overall 1.5 Mbit/s bitrate, but the standard may also work at lower and higher bitrates.

To provide a credible alternative to the analogue video tape recorders, the MPEG-1 solution had not only to provide a video quality at least comparable to the VHS quality but also the special access modes typical of these devices such as fast forward, fast reverse and random access. Thus, the MPEG-1 solution is not only conditioned by coding efficiency but also by another major requirement: the provision of random access functionalities. The consideration of additional requirements depending on the targeted area of use underpins the MPEG standards' evolution which kept adding provisions for new functionalities such as random access, error resilience, low delay, object-based interaction, scalability, metadata, rights protection, etc as needs arose.

As mentioned above, MPEG develops audio and video coding standards in parallel, together with the multiplexing and synchronization specifications in order to provide a complete solution exploiting the available synergies. However, and although designed to be used together, the various specifications correspond to separate parts of the overall standards in order they can also be used independently and with other (also non-MPEG) tools, e.g. a different video coding solution together with the MPEG-1 Systems and Audio solutions.

For this purpose, the MPEG standards are structured in parts (formally speaking independent standards), each one defining a major piece of technology which may be used standalone. MPEG-1 is the MPEG standard with the smallest number of parts, notably five:

- **Systems** (part 1): This part defines the multiplexing of one or more media streams (MPEG-1 Video, Audio or other) with timing information, to form a single stream.
- **Video** (part 2): This part defines a coding format for progressive video (stream and the corresponding decoding process), still largely used nowadays. The target operational environment was storage media at a continuous transfer rate of about 1.5 Mbit/s, but the coding format is rather generic and can be used more widely. This format supports special access functionalities such as fast forward, fast reverse and random access into the coded bitstream. The adopted coding architecture is a hybrid coding scheme based on block-based Discrete Cosine Transform (DCT) applied to a single picture or to a prediction error obtained after temporal prediction (based on one - past - or two pictures – past and future) with motion compensation. DCT is followed by quantization, zigzag scanning and variable length coding. This coding architecture will also be the basis for the MPEG-2 and MPEG-4 video coding formats.
- **Audio** (part 3): This part defines an audio coding format (stream and the corresponding decoding process) for monophonic (32 to 192 kbit/s) and stereophonic (128 to 384 kbit/s) sound. This format includes three hierarchical coding layers – I, II, and III – which are associated to growing complexity, delay and efficiency. MPEG-1 Audio coding solutions are designed for generic audio, and deeply exploit the perceptual characteristics and limitations of the human auditory system, targeting the removal of perceptually irrelevant data. One of these codecs, Layer III, is more commonly known as MP3. Because MP3 made music exchange, downloading, storage, and streaming much easier, piracy and consequently content protection became very quickly major issues. As a side effect, MP3 had a huge impact on the evolution of important multimedia communication technologies, notably peer-to-peer networking, and digital rights management (DRM).
- **Conformance Testing** (part 4): This part defines tests to check if bitstreams (content) and decoders are correct according to the Systems, Video, Audio specifications.
- **Software Simulation** (part 5): This part includes software implementing the tools specified in parts 1, 2 and 3.

Conformance Testing and Software Simulation are present in all MPEG standards and are considered essential for the deployment of MPEG standards. While Conformance Testing targets the provision of checking tests which allow manufacturers to be sure that their products conform according to the standard (and thus should interoperate), Software Simulation provides an implementation which may serve the industry as starting point for the development of compliant products, shortening the time to the market. Both these parts are specific of MPEG standards.

Following the principle that MPEG standards must specify the minimum necessary, MPEG-1 (and following MPEG coding standards) only specifies the coding format and its decoding but not the coding process, leaving much freedom to application developers for competition and improvement since encoders are those that mainly set the coding performance of a codec.

The MPEG-1 standard is still a very popular format nowadays. Needless to say, MP3 is not only one of the most used standards in the multimedia communications arena but also the major responsible for the current revolution in the music industry and business.

III. MPEG-2: GENERIC CODING OF MOVING PICTURES AND ASSOCIATED AUDIO

The continuous developments in digital coding and the growing appetite for digital multimedia solutions set the basis for the next evolution in the MPEG family: a new audiovisual coding standard targeting a wider range of bitrates, more choice in video resolution and support for interlaced video signals.

For the first time, digital convergence played a role by bringing together the coding experts of ITU-T and ISO/IEC (through MPEG) to address the requirements for a common video representation solution in the area of audiovisual entertainment, both broadcasting and storage.

The MPEG-2 standard [4][5] defines a new audiovisual coding solution, mainly addressing digital television and storage at medium and high qualities (including HDTV). MPEG-2 Video became the first MPEG joint specification, published as ISO/IEC 13818 Part 2, ("MPEG-2 Video"), and simultaneously as recommendation ITU-T H.262. Considering the short time elapsed and because many requirements were similar, the MPEG-2 Systems, Video and Audio specifications are largely based on the corresponding MPEG-1 specifications. The most relevant parts of MPEG-2 are:

- **Systems** (part 1): This part deals with the same requirements as MPEG-1 Systems but adds support for 1) error prone environments such as broadcasting; 2) hardware-oriented processing and not only software oriented processing; 3) carrying multiple programs simultaneously without a common time base; and 4) transmission in ATM environments. The fulfillment of these requirements resulted in the definition of two types of MPEG-2 Systems streams: the Program Stream (PS) similar to, and compatible with, MPEG-1 Systems streams, and the new Transport Stream (TS) to carry multiple independent programs.
- **Video** (part 2): This part defines a generic video coding format (stream and the corresponding decoding process) for progressive and interlaced video sequences up to HDTV resolutions. The basic coding architecture is the same as for MPEG-1 Video, adding support for scalable (hierarchical) coding formats, e.g. temporal, spatial, and SNR (Signal to Noise Ratio). Because MPEG-2 Video includes a large set of tools, some applications don't need some of these tools for their utility; in fact, some of these tools would represent an insurmountable burden for the usage of MPEG-2 Video in some applications if a compliant decoder would have to implement all the tools. To be able to provide coding solutions with appropriate complexity for various application scenarios, MPEG-2 Video defines the so-called profiles and levels. Profiles are tool subsets that address the needs of a specific class of applications, while levels are defined for each profile, to limit the memory and computational requirements of a decoder compliant implementation. MPEG-2 Video provides forward compatibility with MPEG-1 Video, meaning that a MPEG-2 Video decoder is able to decode MPEG-1 Video streams. Backward compatibility, meaning that (subsets of) MPEG-2 Video are decodable by MPEG-1 Video decoders, is only provided for specific profiles through scalability.
- **Audio** (part 3): This part defines an audio coding format (stream and the corresponding decoding process) for multichannel audio. Part 3 is also known as "backward compatible (BC)" audio since it provides backward and forward compatibility with MPEG-1 Audio. Because of backward compatibility (an MPEG-1 stereo decoder should be able to create a meaningful version of the original multichannel MPEG-2 Audio stream), MPEG-2 Audio is technically similar to MPEG-1 Audio, which means that it also defines three codecs with growing complexity and performance by means of the same three coding layers. The most important MPEG-2 (BC) Audio added functionality is the support for multichannel audio, typically in 5+1 combinations.
- **Conformance Testing** (part 4): As for MPEG-1, this part defines tests allowing checking if bitstreams (content) and decoders are compliant to the specifications in parts 1, 2 and 3. For video streams, conformance is defined for the various profile@level combinations.
- **Software Simulation** (part 5): As for MPEG-1, this part includes software implementing the tools specified in parts 1, 2 and 3.
- **Digital Storage Media – Command and Control (DSM-CC)** (part 6): Because MPEG-2 also intends to address applications like video on demand where server-client interaction is essential, this part defines generic control commands independent of the DSM type. With these commands, MPEG applications may access local or remote DSMs to perform functions

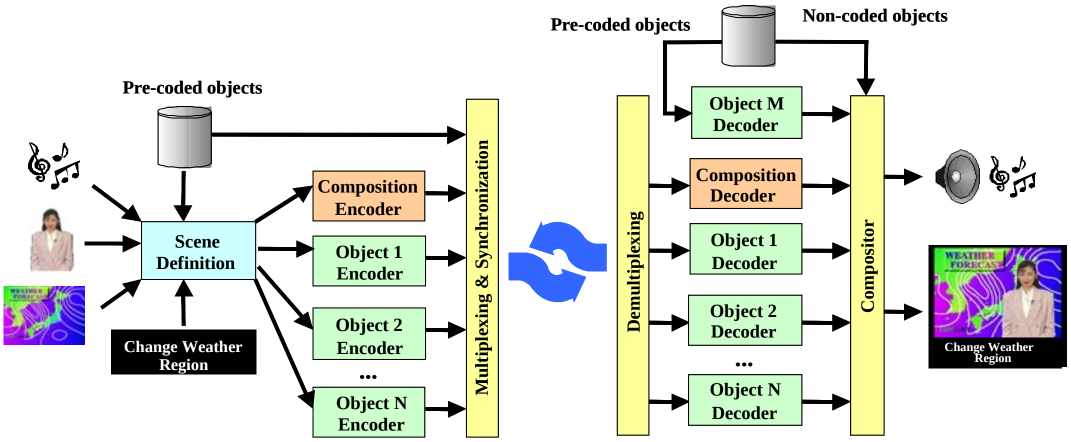

Figure 1: Simplified MPEG-4 coding architecture

specific to MPEG streams without having to know details about the DSMs. These commands apply to MPEG-1 Systems, and MPEG-2 Program and Transport streams.

- **Advanced Audio Coding (AAC)** (part 7): This part (also known as Non Backward Compatible, NBC) defines a multichannel audio coding format, not providing MPEG-1 backward compatibility, and achieving similar qualities at much lower bitrates than MPEG-2 BC Audio. The need for this second MPEG-2 Audio part resulted from the fact the MPEG-2 BC Audio had to compromise quite significantly in terms of coding efficiency to fulfill the MPEG-1 backward compatibility requirement.
- **IPMP on MPEG-2 Systems** (part 11): This part is a recent addition to MPEG-2 specifying further Systems tools that allows the IPMP capabilities developed in the context of MPEG-4 to be used with MPEG-2 Systems streams.

The MPEG-2 standard is very likely the most successful multimedia communications standard in the market since MPEG-2 technology is used in the DVDs and the set top boxes for all the digital television systems in the world, notably DVB, ATSC and ISDB. Digital multimedia entertainment is currently mostly a MPEG-2 arena.

In terms of evolution, it is worthwhile to mention that the biggest conceptual novelty of MPEG-2, scalable video coding, has never been adopted by the industry (especially due to the associated compression efficiency losses regarding non-scalable solutions). Interestingly, in 2006, scalable video coding is again a major activity in MPEG, still mainly targeting to close the coding efficiency gap.

IV. MPEG-4: CODING OF AUDIO-VISUAL OBJECTS

The MPEG-4 standard [6][7], launched by MPEG in 1994, incorporates a major conceptual advance in audiovisual content representation: the object-based representation model. The object-based model avoids the blindness of the frame-based model, which has its roots in analogue television and was adopted by the MPEG-1 and MPEG-2 standards. The new model recognizes that audiovisual content aims at reproduction a world which is made of elements, called the objects. With the adoption of the object-based model, MPEG-4 launches a new approach to multimedia content representation where the audiovisual scene is taken as a composition of independent objects with their own coding, features and behaviors. Figure 1 shows a simplified object-based audiovisual coding architecture where the composition information puts together in the scene a number of audio and visual objects, which are independently accessible since they were independently coded. This architecture provides a full range of interaction capabilities, automatic or user driven.

The novel object-based representation approach brings some major benefits, notably:

1. **Hybrid natural and synthetic coding:** For the first time in multimedia representation, natural and synthetic content have the same status, this means audio and visual objects in the scene can be of any origin, natural or synthetic, e.g. text, speech, frame-based video, arbitrarily shaped video objects, 3D models, synthetic audio, and 2D meshes. It is worthwhile to mention that natural-synthetic (hybrid) content mixes are a growing trend in multimedia content production, e.g. in television and Internet (see Figure 2).

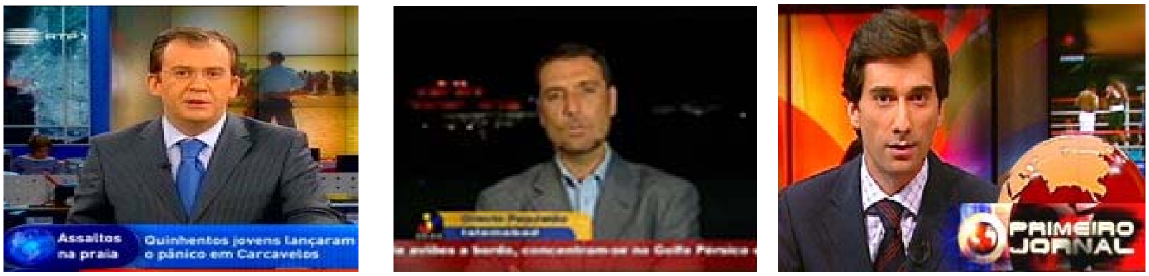

Figure 2: Combining natural and synthetic objects

Content-based interaction and reusing: It is possible to directly interact with the various objects in the scene, changing their properties or behavior, since in object-based scenes the objects are independently represented from each other, and thus independently accessible. This also means that objects may be reused from a scene to another which is an essential feature to decrease costs in content production. Internet developments had already shown by this time that users wanted in the audiovisual world similar interaction capabilities to those available in terms of text and graphics.

2. **Content-based coding:** Because of the object individual representations, it is natural to code each type of object taking benefit of its intrinsic features: this means that a text object will be coded using a text coding tool while a 3D object will be coded using a 3D coding tool. Independently of the interaction capabilities, this approach brings added coding efficiency since the right tools are used for each type of data.

3. **Universal access:** The intrinsic representation flexibility and granularity of the object-based approach fits the needs of mobile and wireless terminals, where access to audiovisual content from anywhere, at anytime has become a major requirement. This flexibility in terms of coding, error resilience and scalability provides the ideal conditions to create adequate content variations for each consumption conditions, targeting the best multimedia experience.

It is important to highlight that the object-based representation model does not specifically target any application scenario or bitrate but brings benefits in a rather horizontal way. This justifies the fact that MPEG-4 will cover a very wide range of bitrates from low bitrate personal mobile communications to high quality studio production.

The wide set of requirements addressed by MPEG-4 led to the specification of about 20 parts in this standard, among them:

- **Systems** (part 1): This part defines the systems architecture and the tools associated with scene description - both the BInary Format for Scenes (BIFS) and eXtensible MPEG-4 Textual (XMT) formats, multiplexing, synchronization, buffer management, and management and protection of intellectual property. It specifies also the MP4 file format designed to be independent of any particular delivery protocol while enabling efficient support for delivery in general. Finally, it specifies MPEG-J which is a Java application engine defining how applications may be contained in a bitstream and executed at the client terminal. The scene description tools are associated to the major conceptual novelty in MPEG-4, this means the object-based representation model since besides coding each object there is now a need to code its behavior and the composition data.
- **Visual** (part 2): This part defines all the coding tools associated to visual objects, both of natural (including arbitrarily shaped video objects) and synthetic origin. While previous MPEG video coding solutions were specified in parts named 'Video', the new naming ('Visual') stresses the fact that from now on natural and synthetic content are at the same level in MPEG standards.
- **Audio** (part 3): This part defines all the coding tools associated to aural objects, both of natural and synthetic origin. Together, the Visual and Audio parts define a large set of coding solutions different not only in terms of functionalities for a certain type of data but also different in terms of the type of data targeted, making MPEG-4 a huge coding tool box. Again profiles and levels allow defining adequate solutions for each application scenario in terms of functionality, efficiency and complexity.
- **Conformance Testing** (part 4): As for previous standards, this part defines tests allowing checking if bitstreams and decoders are correct according to the specifications in the parts defining the technologies. For visual and audio streams, conformance is specified for profile@level combinations where a profile is defined as a set of object types [7].
- **Reference Software** (part 5): This part includes software corresponding to most parts of MPEG-4, notably visual and audio encoders and decoders; this software is copyright free (not patents' free) for compliant products. Unlike in MPEG-1 and MPEG-2, MPEG-4 reference software for

decoders is considered normative, meaning that it has the same status as the textual specification.

- **Delivery Multimedia Integration Framework (DMIF)** (part 6): This part defines a delivery media independent representation format to transparently cross the borders of different delivery environments.
- **Advanced Video Coding (AVC)** (part 10): This part defines a novel frame-based video coding solution providing up to 50% higher coding efficiency than the best video coding profile in MPEG-4 Visual (part 2), for a wide range of bitrates and video resolutions, at the cost of increased complexity. This part has been developed by the Joint Video Team (JVT) created to formalize the collaboration between MPEG and the ITU-T Video Coding Experts Group (VCEG) for the joint development of this standard. MPEG-4 AVC is known within ITU-T as Recommendation H.264. In 2006, the most important MPEG video coding activities - SVC targeting the specification of an efficient scalable video coding solution and MVC targeting the specification of an efficient multiview video coding solution - are related to this MPEG-4 part since both SVC and MVC build on AVC in a backward compatible way. This means both SVC and MVC will be defined as amendments (extensions) to MPEG-4 part 10.
- **ISO Base Media File Format** (part 12): This part defines the ISO base media file format, which is a general format forming the basis for a number of other more specific file formats, notably the MPEG-4, AVC and SVC file formats.
- **Intellectual Property Management and Protection (IPMP) Extensions** (Part 13): This part defines tools to manage and protect intellectual property on audiovisual content and algorithms, so that only authorized users have access to it. This part is associated to another major novelty in MPEG-4 this means the recognition of the importance to provide tools for the management and protection of intellectual property. This trend will reach its apogee in MPEG-21.
- **MP4 File Format** (part 14): This part defines the MP4 file format as an instance of the ISO base media file format (part 12). This was previously included in part 1, but for easier referencing a separate part was created.
- **AVC File Format (**part 15): This part defines a storage format for AVC compressed video streams. This format is based on the ISO base media file format (part 12), which is also used by the MPEG-4, Motion JPEG 2000, and 3GPP file formats, among others.
- **Animation Framework eXtension (AFX)** (part 16): This part defines tools for interactive 3D content operating at the geometry, modeling and biomechanical levels. AFX provides a 3D framework offering advanced features such as compression, streaming, and seamless integration with other audiovisual media, and allowing building high quality creative cross media applications.
- **Lightweight Application Scene Representation (LaSer)** (part 20): This part defines a new scene representation solution targeting a trade-off between expressivity, compression efficiency, decoding and rendering efficiency, and memory footprint.

MPEG-4 represents a large conceptual step forward regarding MPEG-1 and MPEG-2. Since it includes a large amount of tools, it is natural that not all of them reached the same degree of success. In particular, MPEG-4 was considerably affected by the significant delay in setting licensing conditions for some of its technology. Setting licensing conditions is an exercise made outside MPEG by the companies owning the relevant patents and MPEG has no influence on the speed and results of this exercise.

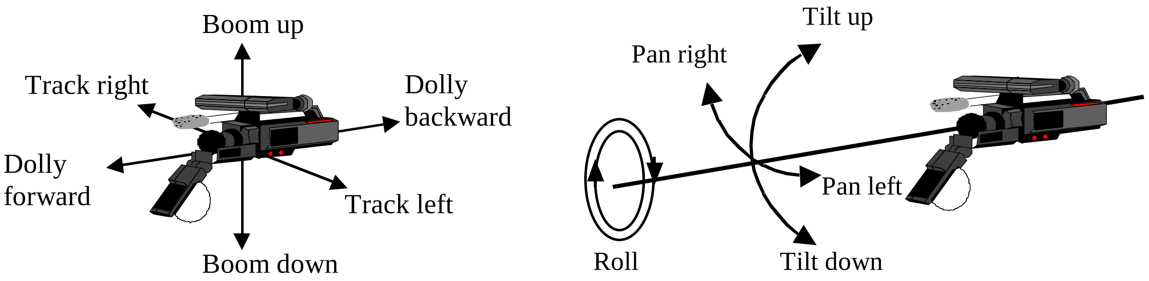

Figure 3: Types of MPEG-7 camera motion

V. MPEG-7: MULTIMEDIA CONTENT DESCRIPTION INTERFACE

It is clear that the deployment of the MPEG-1, MPEG-2 and MPEG-4 standards has made much easier to acquire, produce and distribute audiovisual content. However, this easiness and associated content profusion creates a huge content management challenge since the more content there is, the harder it is to manage, retrieve and filter. Since content has value only if it can be retrieved and filtered, quickly and efficiently, MPEG recognized, around 1996, that after addressing the content coding problem it was time to address the multimedia content management problem. A major tool for the efficient management of multimedia data is the availability of the so-called metadata or ‘data about the data’, which represents the data at a different level, targeting retrieval and filtering and not anymore visualization and hearing.

Thus, following a natural evolutional process, MPEG launched, in 1996, the MPEG-7 project [8][9], formally called ‘Multimedia Content Description Interface’ with the goal to specify a standard way of describing various types of audiovisual information such as elementary pieces, complete works and repositories, irrespective of their representation format or storage medium.

As for past MPEG standards, MPEG-7 is generic and thus provides content description solutions for a large set of application domains, which are also media and format independent, object-based, and extensible. MPEG-7 description tools are able to operate at different levels of abstraction, from low-level, automatic and often statistical features, to high-level features conveying semantic meaning. Thus MPEG-7 offers the possibility to combine in a single description low-level and high-level features which is a unique feature of MPEG-7. To reach its purposes, MPEG-7 defines two major types of tools: 1) descriptors which are a representation of a feature defining the syntax and the semantics of the feature representation where a feature is a “distinctive characteristic of the data that signifies something to somebody” (examples are a time-code for representing duration, color moments and histograms for representing color, and a character string for representing a title); 2) description schemes (DS) which specify the structure and semantics of the relationships between its components, which may be both descriptors and description schemes (a simple example is a description of a movie, temporally structured as scenes and shots, including some textual descriptors at the scene level, and color, motion, and audio amplitude descriptors at the shot level). MPEG-7 descriptions may be expressed by textual streams, using the so called Description Definition Language (DDL), and by binary streams, using the Binary format for MPEG-7 data (BiM), which is basically a DDL compression tool.

The MPEG-7 standard defines eleven parts, the most relevant of which are:

- **Systems** (part 1): This part defines the tools for: 1) transporting and storing MPEG-7 descriptions in an efficient way using the BiM representation format (note that the BiM can be seen as a general XML compression tool); 2) synchronizing MPEG-7 descriptions with the content they describe (MPEG-7 descriptions may be delivered independently or together with the content they describe); and 3) managing and protecting the intellectual property associated with the descriptions.
- **Description Definition Language (DDL)** (part 2): This part defines a language for creating new description schemes as well as extending and modifying existing ones. The DDL is based on W3C’s XML (eXtensible Markup Language) Schema Language; some extensions to XML Schema were developed in order to address all the identified DDL requirements.

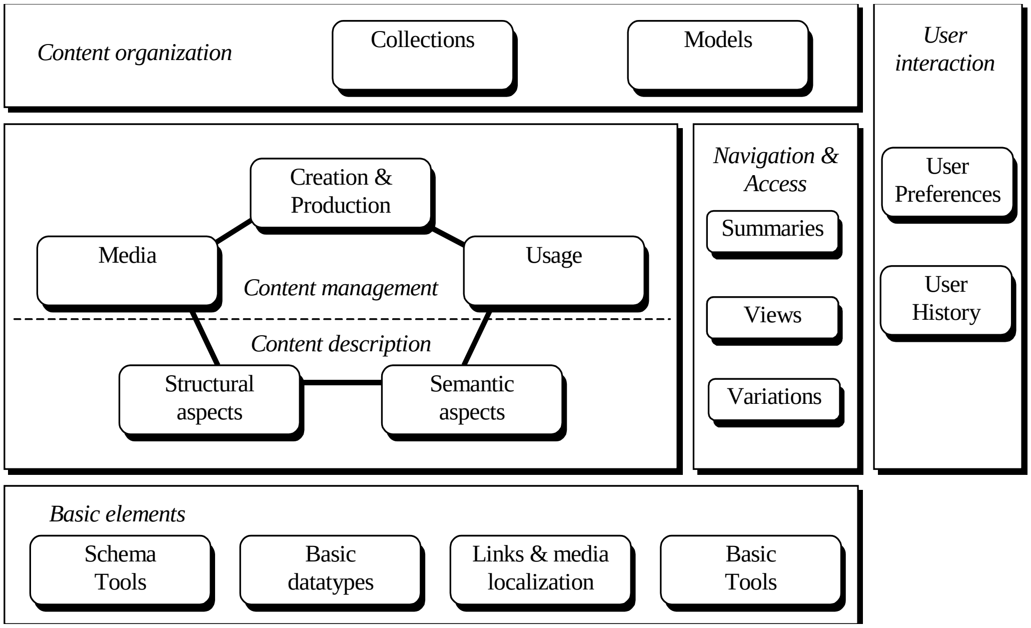

Figure 4: Overview of the MDS description tools

- **Visual** (part 3): This part defines the visual description tools, notably basic structures and descriptors or description schemes, for the description of visual features and for the localization of the described objects in the image or video sequence. The MPEG-7 visual descriptors cover five basic visual features: color, texture, shape, motion (see Figure 3), and localization (also a face recognition descriptor is defined).
- **Audio** (part 4): This part defines the audio description tools which are organized in areas such as timbre, melody, silence, spoken content, and sound effects.
- **Multimedia Description Schemes (MDS)** (part 5): This part defines the description tools dealing with generic as well as multimedia entities (not visual or audio specific). MDS description tools can be grouped into six different classes according to their functionality (see Figure 4): 1) content description: structural, and semantic aspects; 2) content management: media, usage, creation and production; 3) content organization: collections and models; 4) navigation and access: summaries, variations and views; 5) user: user preferences, and usage history; and 6) basic elements: datatype and structures, schema tools, link & media localization, and basic DSs.
- **Reference Software** (part 6): Again, this part includes software implementing the tools specified in the other parts. As for MPEG-4, this software is normative and can be used free of copyright for implementing products compliant to the standard.
- **Conformance Testing** (part 7); Again, this part defines procedures allowing to check if description streams are according to the specifications in the other parts; conformance for descriptions is specified for a profile@level combination.
- **Profiles and Levels** (part 9): This part defines description profiles and levels. A description profile defines a subset of all the description tools available in MPEG-7, supporting a set of functionalities; a level of a description profile defines further constraints on conforming descriptions, constraining their maximum complexity. MPEG-7 is the only MPEG standard with a separate part for defining profiles and levels because description profiles include tools from various MPEG-7 parts; in the other MPEG standards, profiles and levels are defined in the corresponding parts, e.g. video profiles in the Video part and audio profiles in the Audio part.

While MPEG-7 defines a rather rich, and powerful set of multimedia content description tools without competition in other metadata standards, its deployment is still minor. Possible reasons for this lack of success may be the lack of licensing conditions, the still little usage of low-level descriptors, and the still unclear meaning and value of metadata interoperability for many application scenarios.

With MPEG-7, MPEG puts another important brick in its multimedia technology building; among MPEG-1, MPEG-2, MPEG-4 and MPEG-7, MPEG standards provide high performance and functionality rich solutions for multimedia content coding and description. It is than more than natural that next MPEG standards try to open rather new doors in the world of multimedia technology.

VI. MPEG-21: MULTIMEDIA FRAMEWORK

After defining powerful standard solutions for multimedia content coding and description, MPEG was expecting the digital multimedia business to burst but that did not happen to the extent expected. In 2000, MPEG discussed again the multimedia landscape and the deployment of its standards, and acknowledged that the widespread deployment of multimedia applications required more than a loose collection of standards. Multimedia consumption and commerce remained non-transparent and were not happening in a large scale. In trying to answer questions such as ‘Do all existing multimedia-related standard specifications fit together?’, ‘Does anybody know how they fit together?’, ‘Are there specifications for all the necessary technical elements for multimedia transactions?’, ‘Which standard activities are most relevant ?’, and ‘Who is making the "glue" that will allow standards to fit together?’, MPEG concluded that there was a need to address the multimedia problem at a higher level and to consider the complete multimedia consumption chain.

As a result, MPEG decided to develop the MPEG-21 standard, formally called “Multimedia framework” [10][11]. In terms of vision, the main objective of MPEG-21 is to enable the transparent and augmented use of multimedia resources across a wide range of networks, devices, and communities. A key assumption is that every human is potentially an element of a network involving billions of content providers, value adders, packagers, service providers, consumers, and resellers. This means that besides client-server-based applications, peer-2-peer networking and the resulting flexibility of user roles have been part of MPEG-21 thinking since the early days.

The overall MPEG-21’s goal is to create an interoperable multimedia framework by: 1) defining the "big picture" to understand how the components of the framework are related and identify where gaps in the framework exist; 2) filling the gaps by developing new standard specifications where needed with the involvement of other bodies, where appropriate; and 3) integrating the standard tools to support harmonized technologies for the management of multimedia content. This goal makes the MPEG-21 standard a top-down standard in opposition to previous MPEG standards which were clearly bottom-up. This ‘filling the gaps’ approach also makes the MPEG-21 standard a bit miscellaneous as a result of its purpose to provide standard solutions where they are missing in the multimedia framework. Of course, previous MPEG standards are examples of technologies that should fit in the MPEG-21 framework and thus for which no specific MPEG-21 action is needed.

Since MPEG fully acknowledged the size of the MPEG-21 challenge and the fact that it only had a significant background in standards related to multimedia content delivery, management and representation, collaboration with other bodies was envisioned. The aim was to maximize interoperability, minimize the overlap between concurrent activities, and share common technologies.

The basic MPEG-21 concepts relate to the ‘What’ and ‘Who’ within the multimedia framework. In this context, the ‘What’ is a Digital Item which is a structured digital object with a standard representation, identification and metadata within the MPEG-21 framework. The ‘Who’ is a User (with a capital U) that interacts in the MPEG-21 environment or makes use of a Digital Item, including individuals, consumers, communities, organizations, corporations, consortia, governments and other standards bodies and initiatives around the world. The User roles include creators, consumers, rights holders, content providers, distributors, etc, which means that in MPEG-21 there is no major technical distinction between providers and consumers. Each User in MPEG-21 assumes specific rights and responsibilities according to their interaction with other Users. Because it was considered a major limitation for the growth of the digital multimedia world, it is also a major MPEG-21 requirement that Users are able to express and manage their interests, e.g. rights, in Digital Items.

In practice, a Digital Item is a combination of resources, metadata, and structure (see Figure 5). The resources are the individual assets or (distributed) content. The metadata describes (distributed) data about or pertaining to the Digital Item as a whole or also to the individual resources in the Digital Item. Finally, the structure relates to the relationships among the parts of the Digital Item, both resources and metadata. An example of a Digital Item is a music compilation including the music but also photos, videos, animation graphics, lyrics, scores, MIDI files, interviews with the singers, news related to the songs, statements by an opinion maker, ratings of an agency, position in the hit list, navigational information driven by user preferences, bargains, etc. Notice that the notion of Digital Item is much closer

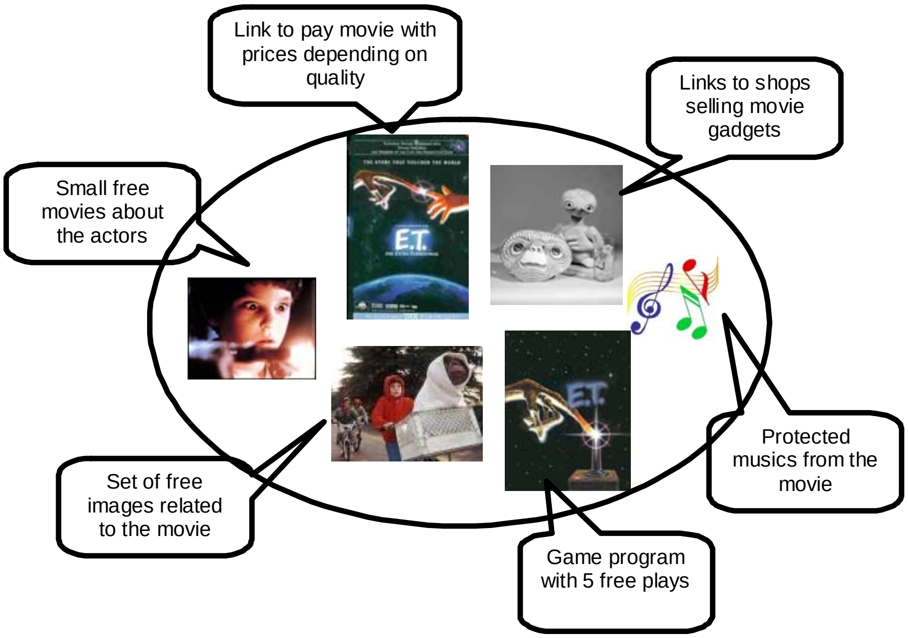
Link to pay movie with prices depending on quality
Links to shops selling movie gadgets
Small free movies about the actors
Protected musics from the movie
Set of free images related to the movie
Game program with 5 free plays

Fig. 5: Example of Digital Item related to the movie "ET"

to the real world, this means to what people buy in the shops, than the technologically driven notions of MPEG-4 and MPEG-7 elementary streams which leave a gap from the technology to the applications. The Digital Item is thus the fundamental unit for distribution and transaction within the MPEG-21 framework.

In conclusion, MPEG-21 provides a framework in which Users interact and the object of the interaction is a Digital Item. It is important to notice that MPEG-21 does not set a rigid framework architecture and associated set of technologies but rather as many architectural solutions as needed by shaping the framework to the individual needs of each application; for example, if no rights management is needed, than a framework without rights related tools is targeted.

As usual, the MPEG-21 requirements were addressed by specifying technology in various parts, notably:

- **Vision, Technologies and Strategy** (part 1): This part defines the multimedia framework and its architectural elements together with the functional requirements for their specification.
- **Digital Item Declaration (DID)** (part 2): This part defines a uniform and flexible abstraction and interoperable schema for declaring Digital Items. These declarations provide the structure for Digital Items and thus for complex multimedia assets.
- **Digital Item Identification (DII)** (part 3): This part defines the framework for the identification of any entity regardless of its nature, type or granularity.
- **Intellectual Property Management and Protection (IPMP)** (part 4): This part defines the means to enable content to be persistently and reliably managed and protected across networks and devices.
- **Rights Data Dictionary (RDD)** (part 5): This part defines a dictionary of key terms which are required to describe rights of all Users.
- **Rights Expression Language (REL)** (part 6): This part defines a machine-readable language that allows to declare rights and permissions using the terms as defined in the Rights Data Dictionary.
- **Digital Item Adaptation (DIA)** (part 7): This part defines description tools for usage environment and content format features; these descriptions are instrumental to provide the user with the most adequate multimedia content, depending on the relevant terminal, network, user preferences and the natural environment where users are consuming the content.
- **Reference Software** (part 8): This part includes software implementing the tools specified in the other MPEG-21 parts; again, this software can be freely used in MPEG-21 compliant products.
- **File Format** (part 9): This part defines a file format for the storage and distribution of Digital Items.
- **Digital Item Processing** (part 10): This part defines mechanisms that provide for standardized and interoperable processing of the information in Digital Items.
- **Conformance** (part 14): Again, this part defines conformance testing for other parts of MPEG-21.

- **Event Reporting** (part 15): This part defines the syntax and semantics of a language to express Event Report Requests and Event Reports; for example, this enables Users within the multimedia framework to monitor the use of Digital Items, and to monitor the load of networks.
- **Binary Format** (part 16): This part defines a a binary format (based on MPEG-7 Systems tools), which allows the binarization, compression, and streaming of some or all parts of a Digital Item.
- **Digital Item Steaming (DIS)** (part 18): This part defines technology for the incremental delivery of a DI (DID, metadata, resources) in a piece-wise fashion with temporal constraints such that a receiving Peer may incrementally consume the DI.

From above, it is rather evident that Digital Items play a central role in MPEG-21; this concept makes a big step towards dealing in MPEG standards with content entities which are close to real world multimedia entities such as a (legally) downloaded music with its metadata or a DVD including several multimedia pieces. It is also evident that MPEG-21 is much centered on intellectual property management and protection, and thus digital rights management, since these are nowadays key technologies for the success of many multimedia business models. MPEG-21 aims to guarantee a higher degree of interoperability by focusing also on how the various elements of a multimedia application infrastructure should integrate and interact.

VII. MPEG-A: MULTIMEDIA APPLICATION FORMATS

It is well known that the major purpose of specifying standards is to provide interoperability, which refers to the ability of a system, or a product, to work with other systems or products without special effort required by the user. In MPEG, the user has always been at the center of all decisions since ‘happy users’ should result in ‘happy (and wealthy) industries’.

In the context of the MPEG-1, MPEG-2, MPEG-4, MPEG-7 and MPEG-21 standards, MPEG tried to provide a trade-off between normative specifications and industry choice, leaving the manufacturers some flexibility, for example in terms of the way some MPEG technologies may be combined together.

While MPEG has defined profiles and levels for various parts of its standards, e.g. MPEG-2 Video, MPEG-4 Audio and Visual, with the objective to provide interoperable solutions at reasonable complexity for certain classes of applications, it has never defined any standard combinations of tools or profiles across different standards or parts of standards. In the area of digital television, for example, MPEG never specified a combination of a MPEG-2 Video profile@level with a MPEG-2 Audio layer, leaving the industry or certain industry fora (e.g. DVB, ATSC) the role to make these choices.

However, the growing number of tools in the standards, and also the number of profiles and levels, has made increasingly difficult for industries to select combinations of tools or profiles. Moreover, with time, and following the evolution of multimedia applications, MPEG decided to provide a range of tools which is not only related to media coding but also to metadata, digital rights management, content adaptation, etc. This enlarged range of technical solutions makes even more difficult to select the right combination of tools, especially by someone who was not involved in the development process. It is also more and more common that users of MPEG standards are not familiar with the details of all MPEG standards and parts of standards, while at the same time they need complete solutions addressing a specific application and not just standard solutions targeting a specific media, e.g. video coding. This situation has led not only to different industry consortia in related application domains picking different solutions, but also to users using proprietary ‘complete solutions’. Both cases result in reduced interoperability, which goes against the essential MPEG objectives.

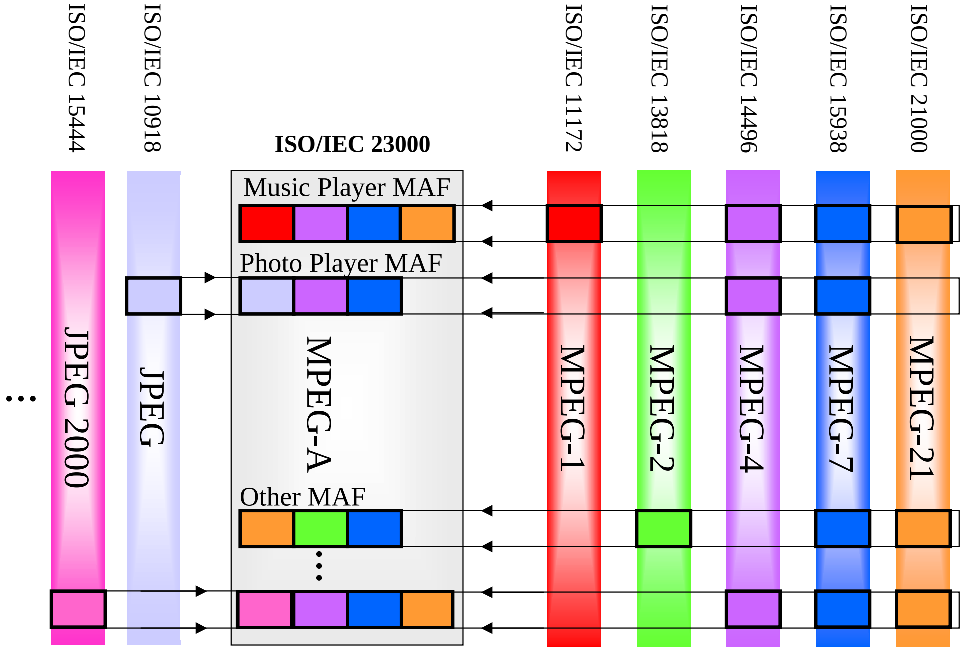

Figure 6: Conceptual overview of MAFs

To address this situation and increase the deployment of its standards, MPEG decided, in 2004, to launch a new standard formally known as ISO/IEC 23000, "Multimedia Application Formats", and also called MPEG-A [12][13]. The MPEG-A standard targets the definition of Multimedia Application Formats (MAFs) which are basically 'super-formats' combining tools defined across existing MPEG standards or parts of standards (see Fig. 6). Whenever needed, these formats may also include non-MPEG tools, e.g. in areas not addressed by MPEG standards. These 'super-formats', e.g. a combination of audio and video coding formats with some metadata, bring the notion of MPEG interoperability to a new dimension. Instead of interoperability associated with a single domain, MPEG-A associates now interoperability with complete application-driven solutions. With these application formats, MPEG takes the responsibility to provide the users with adequate combinations of tools across MPEG standards and parts of standards, not relying anymore on others outside MPEG to make those choices. This objective makes clear that the MPEG-A standard does not target the specification of new MPEG tools but mostly combinations of previously defined tools, ideally in terms of profiles and levels However, if needed, new tools, profiles and levels may be defined and added to the right MPEG standard in order to be used in the relevant MAF.

To reach its objectives, the MPEG-A standard will consist of various parts, each defining one or more related multimedia application formats. For each MAF, the specification will include not only a textual definition but also reference software to ease the adoption of the MAF in question by the industry, including small companies.

In the summer of 2006, one MAF has already been completed: the Music Player MAF. The Music Player MAF addresses music library applications providing an easy way for users to exchange collections of musics, together with associated data and metadata. With this purpose, the Music Player MAF specification defines how to carry MP3 audio along with MPEG-7 metadata within either the MPEG-4 or MPEG-21 File Formats. In addition, JPEG images may also be included, e.g. with the cover of a record or pictures of the musicians. Other MAFs are under development such as the so-called Photo Player MAF and the Protected Music Player MAFs, which add different protection capabilities to the non-protected, already defined, Music Player MAF.

VIII. FINAL REMARKS

In recent years, MPEG standards have played a major role in providing the industry with efficient solutions to solve most multimedia communications problems. After MPEG-21, where a framework and not anymore only tools was targeted, MPEG-A represents a natural evolution of MPEG standards, recognizing that interoperability needs have changed with time and the growing complexity of multimedia applications.

Following the big successes with previous standards, it is time to wait and see if MPEG will provide once more the industry, and through it the users, what they really need.

REFERENCES

[1] http://www.iso.org/iso/en/ISOOnline.frontpage

[2] MPEG Page, http://www.chiariglione.org/mpeg/.

[3] ISO/IEC 11172:1991, “Coding of Moving Pictures and Associated Audio at up to about 1.5 Mbit/s”, 1991.

[4] ISO/IEC 13818:1994, “Generic Coding of Moving Pictures and Associated Audio”, 1994.

[5] B. Haskell, A. Puri, A. Netravali, “Digital Video: an Introduction to MPEG-2”, Chapman & Hall, 1997.

[6] ISO/IEC International Standard 14496:1998, “Coding of Audio-Visual Objects”, 1998.

[7] F. Pereira, T. Ebrahimi (ed.), “The MPEG-4 Book”, Prentice Hall, 2002.

[8] ISO/IEC 15938:2002, “Multimedia Content Description Interface”, 2002.

[9] B. S. Manjunath, P. Salembier and T. Sikora (ed.), “Introduction to MPEG-7: Multimedia Content Description Language”, John Wiley & Sons, 2002.

[10] ISO/IEC 21000:2002, "Multimedia Framework”, 2002.

[11] I. Burnett, F. Pereira, R. Van de Walle, R. Koenen (ed.), “The MPEG-21 Book”, John Wiley & Sons, 2006.

[12] ISO/IEC 23000:2005, "Multimedia Application Formats”, 2005.

[13] K. Diepold, F. Pereira, W. Chang, “MPEG-A: multimedia application formats”, IEEE Multimedia, vol. 12, nº 4, pp. 34 – 41, October-December 2005

Fernando Pereira (S’92–M’93–SM’99) was born in Vermelha, Portugal, in October 1962. He received the Licenciatura, M.Sc., and Ph.D. degrees in electrical and computer engineering from Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Lisboa, Portugal, in 1985, 1988, and 1991, respectively.

He has been participating in the work of ISO/MPEG for many years, notably as the head of the Portuguese delegation, chairman of the MPEG Requirements group, and chairing many ad hoc groups related to the MPEG-4 and MPEG-7 standards. His current areas of interest are video analysis, processing, coding and description, and multimedia interactive services. He is a member of the Scientific and Program Committees of tens of international conferences and workshops. He has contributed more than 180 papers to journals and international conferences. Dr. Pereira is currently a Professor with the Electrical and Computers Engineering Department at IST. He is responsible for the participation of IST in many national and international research projects. He is a member of the Editorial Board and Area Editor on Image/Video Compression of the Signal Processing: Image Communication Journal and an Associate Editor of IEEE Transactions of Circuits and Systems for Video Technology, IEEE Transactions on Image Processing, and IEEE Transactions on Multimedia. He won the 1990 Portuguese IBM Award and an ISO Award for Outstanding Technical Contribution for his participation in the development of the MPEG-4 Visual standard, in October 1998.