MPEG-7 Systems: Overview

Olivier Avaro and Philippe Salembier

***Abstract*—This paper gives an overview of Part 1 of ISO/IEC 15939 (MPEG-7 Systems). It first presents the objectives of the MPEG-7 Systems activity. In the MPEG-1 and MPEG-2 standards, "Systems" referred only to overall architecture, multiplexing, and synchronization. In MPEG-4, in addition to these issues, the Systems part encompasses interactive scene description, content description, and programmability. MPEG-7 brings new challenges to the Systems expertise, such as languages for description representation, binary representation of descriptions, and delivery of descriptions either separate or jointly with the audio–visual content. The paper then presents the description of the MPEG-7 Systems specification, starting from the general architecture up to the description of the individual MPEG-7 Systems tools. Finally, a conclusion describes the status of the standardization effort, as well as future extensions of the specification.**

***Index Terms*—APIs, architecture, audio–visual, binary format, content description, MPEG-7 systems, multiplex, programmability, specification, synchronization, tools.**

I. Introduction

THE concept of "Systems" in MPEG has evolved dramatically since the development of the MPEG-1 and MPEG-2 standards. In the past, "Systems" referred only to overall architecture, multiplexing, and synchronization. In MPEG-4, in addition to these issues, the Systems part encompasses interactive scene description, content description, and programmability.

MPEG-7 brings new challenges to the Systems expertise, such as languages for description representation, binary representation of descriptions, and delivery of descriptions either separate or jointly with the audio–visual content. The combination of the new possibilities of describing audio–visual content offered by MPEG-7 Systems, and the efficient description tools provided by the visual, audio, and MDS parts of the standard promise to be the foundation of a new way of thinking about audio–visual information.

Indeed, before MPEG-7, audio–visual data was mainly an opaque series of bits. Only the decoding of these bits would give some information about what the data was about and what the user could do with it. The decoding process involves, in general, complex and memory-demanding operations and requires high bandwidth in networked environments.

With the use of MPEG-7 descriptors and description schemes [1], MPEG-7 provides a way to get information about the audio–visual data without the need of performing the actual decoding of these data. The MPEG-7 Systems specification completes the picture by linking MPEG-7 description with the audio–visual content and providing an efficient binary representation of the description and description schemes data, in the best MPEG tradition.

This paper gives an overview of MPEG-7 Systems. It is structured around the objectives and architecture of MPEG-7 Systems as follows.

- *Objectives:* This section describes the motivations and the rationale behind the development of the MPEG-7 Systems specifications. As with all MPEG activities, MPEG-7 Systems is guided by a set of requirements [2], i.e., the set of objectives that must be satisfied by the specifications resulting from the work or activities of the sub-group. This paper gives a particular attention to the way the requirements of MPEG-7 Systems are derived from the principal concept behind MPEG-7 *viz.* the Multimedia Content Description Interface.
- *Architecture and Tools:* This section describes the overall structure of MPEG-7, known as the "MPEG-7 Systems Architecture." A complete walkthrough of an MPEG-7 session highlights the different phases that a user will, in general, follow in consuming MPEG-7 descriptions. MPEG-7 is a "toolbox" standard, providing a number of tools, sets of which are particularly suited to certain applications. This section provides as well a functional description of the MPEG-7 Systems tools. They are fully specified in [3] and [4].

In concluding this look at MPEG-7 Systems, this paper provides a description of the status of the standardization effort as well as short descriptions of some Systems technologies that are currently under consideration by the MPEG Standardization Committee for future versions of the MPEG-7 standard.

II. Objectives

A. Requirements

To understand the rationale behind the MPEG-7 activity, a good starting point is one of the most fundamental MPEG-7 documents *viz.* the MPEG-7 Requirements [2]. This document gives an extensive list of the objectives that needed to be satisfied by the MPEG-7 specifications. It also defines the terminology that is used further in this paper.

MPEG-7 Systems requirements may be categorized into two groups: traditional MPEG Systems Requirements as described in Section II-B and specific MPEG-7 Systems requirements detailed in Section II-C.

B. Traditional MPEG Systems Requirements

Key requirements for the development of the Systems specifications in MPEG-1, MPEG-2, and MPEG-4 were to enable the delivery of coded audio, video, and user-defined private data, and to incorporate timing mechanisms to facilitate synchronous

Manuscript received September 2000; revised March 19, 2001.

O. Avaro is with France Telecom R&D DIH/HDM, 35 A Schützenhüttenweg, Germany (e-mail: olivier.avaro@francetelecom.com).

P. Salembier is with the Department of Signal Theory and Communication, Universitat Politecnica de Catalunya, 08034 Barcelona, Spain (e-mail: philippe@gps.tsc.upc.es).

Publisher Item Identifier S 1051-8215(01)04994-1.

decoding and presentation of these data at the client side. These requirements also constitute a part of the fundamental requirements set for MPEG-7 Systems and are further described below.

1) *Delivery:* The multimedia descriptions are to be delivered using a variety of transmission and storage protocols. Some of these delivery protocols include streaming, e.g., live broadcast of the descriptions along with the content. In these cases, the multimedia descriptions have to be transmitted piece by piece, in order to match the delivery of the descriptions to clients with limited network and terminal capabilities. Delivery implies, as well, the definition of multiplexing tools to embed multiple MPEG-7 descriptions into a single data stream or to embed MPEG-7 description(s) into a single data stream together with associated content.
2) *Synchronization:* Typically, the different components of an audio–visual presentation are closely related in time. For some applications, the descriptions information has to be presented to the user at precise instants in time, together with the content (e.g., before, at the same time, or after the content has been displayed). The MPEG-7 representation needs to allow a precise definition of the notion of time so that data received in a streaming manner can be processed and presented at the right instants in time, and be temporally synchronized with each other.
3) *Stream Management:* Finally, the complete management of streams of audio–visual information including MPEG-7 descriptions implies the need for certain mechanisms to allow an application to consume the content. These include mechanisms such as unambiguous location of the data, identification of the data type, description of the dependencies between data elements, association of descriptions with the content (e.g., with a content elementary stream, or part of it), and access to the intellectual property information associated to the data.

C. MPEG-7 Specific Systems Requirements

In addition to these requirements, MPEG-7 brought specific needs to be solved at the Systems level: languages for the representation of description schemes and representation of binary and dynamic descriptions. The first requirement is solved by the MPEG-7 Description Definition Language (MPEG-7 DDL), which is fully discussed in a companion paper in this Special Issue [6]. The requirements on the dynamic and binary descriptions (referred to as the MPEG-7 BiM—Binary format for Metadata) are described below.

It is expected that in some operational MPEG-7 environments, delivery (network or storage) resources will be scarce. Therefore, data need to be compressed and transferred in an incremental way. The main requirements for the BiM are therefore to provide a compact and streamable representation of the MPEG-7 descriptions. The BiM is equivalent to the textual (XML) description defined by the MPEG-7 DDL and there is a bi-directional mapping between the BiM and the textual DDL-based representation.

In addition, it is expected that some applications will use MPEG-7 BiM-encoded content directly, without necessarily an intermediate step of reconstruction of the textual representation. The binary format, therefore, allows fast parsing of the MPEG-7 streams. It is also designed in a way that allows a BiM parser to check the syntactical correctness, e.g., well-formedness, and the validity regarding normative aspects of an MPEG-7 bitstream.

III. Architecture and Tools

A. MPEG-7 Terminal Architecture

The information representation specified in the MPEG-7 standard provides the means to describe multimedia content. The entity that makes use of such representation is generically referred to as a "terminal". This terminal may correspond to a standalone application or be part of an application system. The overall architecture of an MPEG-7 terminal is depicted in Fig. 1.

The transmission/storage medium appears at the bottom of the figure. This medium refers to the lower layers of the delivery infrastructure (network as well as storage layers and below). These layers deliver multiplexed streams to the Delivery layer. The transport of the MPEG-7 descriptions can occur on a variety of delivery systems. These include for example MPEG-2 Transport Streams, Internet Protocol (IP), or MPEG-4 (MP4) files or streams. The delivery layer encompasses mechanisms allowing synchronization, framing and multiplexing of MPEG-7 description. MPEG-7 description may be delivered independently or together with the content they describe. After the demultiplexing step, the output of the delivery layer is a set of elementary streams. These elementary streams provide pieces of information about MPEG-7 description, about the schema that is used to define the description, or even about the multimedia content itself. MPEG-7 is only standardizing streams for the descriptions and the schemas, that are the XML-Schema specifications of the MPEG-7 descriptors and description schemes.

The information transmission does not always involve a pure downstream scenario (that is, from the server to the client). The MPEG-7 terminal architecture allows data to be conveyed back from the terminal to the transmitter or server. This type of data is called upstream data and typically involves queries formulated by the end user or request of specific pieces of information.

The delivery layer also provides MPEG-7 elementary streams to the compression layer. MPEG-7 elementary streams consist in consecutive and individually accessible portions of data named *access units*. An access unit is the smallest data entity to which timing information (e.g., for inter or intra stream synchronization) can be attributed. MPEG-7 elementary streams contain information of different nature.

1) *Schema Streams*: This information defines the structure of the MPEG-7 description. For some applications, one can assume that the application has the knowledge of the schema that has been used to generate the description. In this case, the schema does not need to be transmitted to the terminal;
2) *Description Streams:* This information can be either the complete MPEG-7 description of the multimedia content or fragments of this description. Fragments are used in situations where the transmission of the entire description in a single chunk of data is not appropriate. In this case, the MPEG-7 description has to be reconstructed combining various fragments.

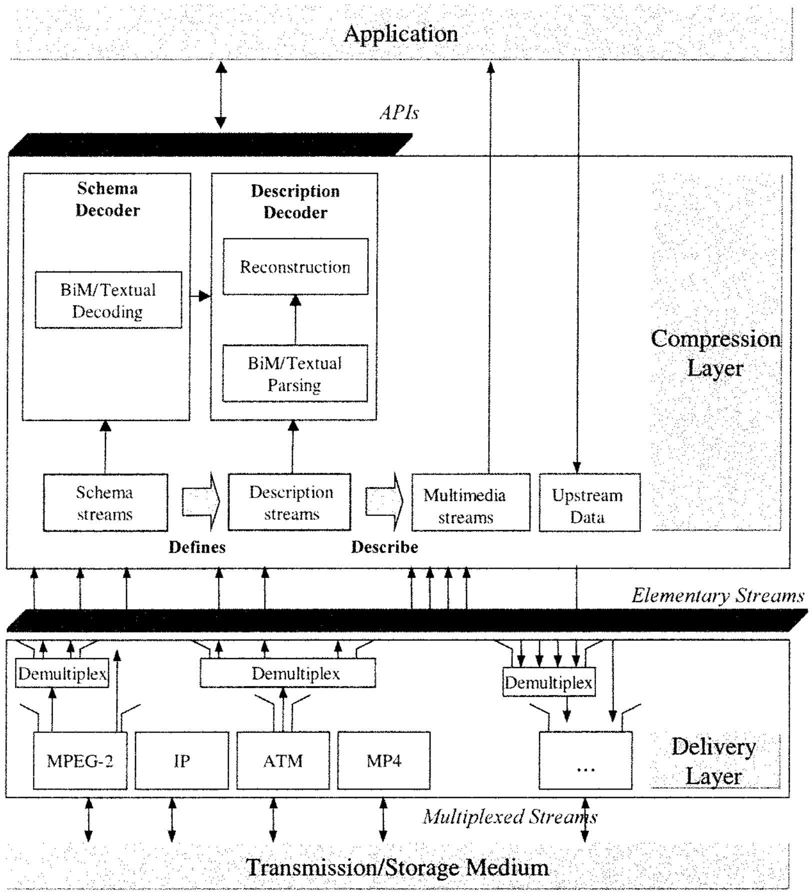

Fig. 1. MPEG-7 Systems architecture.

MPEG-7 description can be represented either in textual format (XML), in binary format (BiM), or a mixture of the two formats, depending on application usage. MPEG-7 defines a unique mapping between the binary format and the textual format. A bi-directional loss-less mapping between the textual representation and the binary representation is possible. Still, it shall not always be used: some applications may not want to transmit all the information contained in the textual representation and may prefer to use a lossy transmission that is more efficient in terms of bandwidth.

At the compression layer, the flow of Access Units (either textual or binary encoded) is parsed, and the MPEG-7 description is reconstructed. MPEG-7 does not mandate the reconstruction of a textual representation as an intermediate step of the decoding process. The MPEG-7 binary stream can be parsed by the BiM parser, transformed in textual format and then processed in textual format for further reconstruction. In an alternative scenario, the binary stream can be parsed by the BiM parser and then forwarded in proprietary format to further processing.

Once reconstructed, the application is then ready to exploit the MPEG-7 descriptions, possibly along with the multimedia elementary streams. The following sections provide more details about the three major MPEG-7 Systems tools: the MPEG-7 access unit, the BiM, and the delivery of the MPEG-7 descriptions.

B. Access Units

MPEG-7 access units are structured as commands encapsulating the MPEG-7 description or the schema. Commands provide the dynamic aspects of the MPEG-7 description: they allow a description to be delivered in a single chunk or to be fragmented in small pieces. This feature is illustrated in Fig. 2. The MPEG-7 description or schema in textual (XML) format is physically a tree structure. The upper part of Fig. 2 illustrates how this tree can be encapsulated in a single Access unit that is transmitted to the terminal. The second scenario is illustrated in the lower part of Fig. 2. In this case, the MPEG-7 description is fragmented in three pieces that are encapsulated in different Access units. The final description is reconstructed by “adding” the content of access units 2 and 3 to the appropriate tree node of the content of access unit 1. Beside the “add” functionality, commands also allow basic operations on the MPEG-7 description such as updating a descriptor value, deleting part of the description or adding new schema. The reconstruction stage of the compression layer updates the description information and associated schema information by consuming these commands. In summary, the access units define: 1) the type of command to be performed in the terminal (add, delete, update, etc.); 2) the MPEG-7 description or schema to be used for the modification; and 3) the location of the tree node where this modification has to be done.

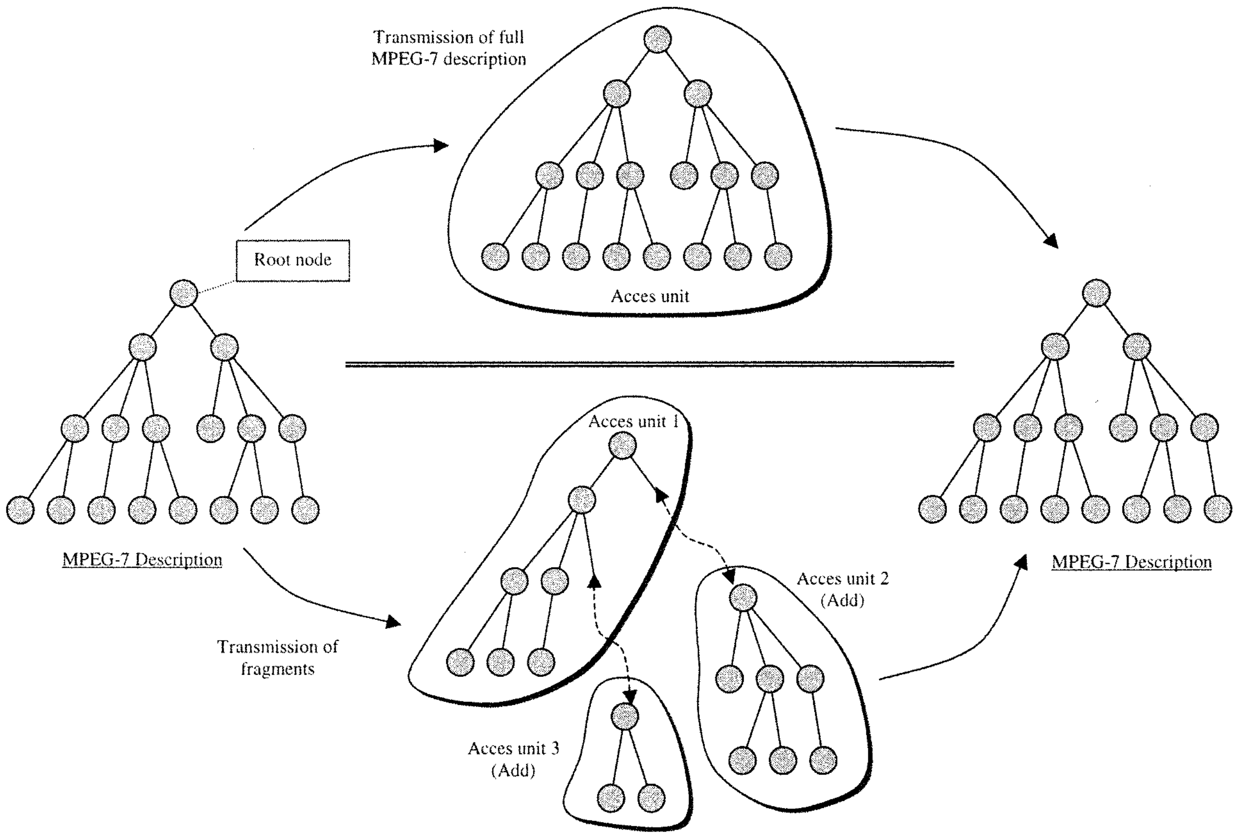

Fig. 2. MPEG-7 description and access units.

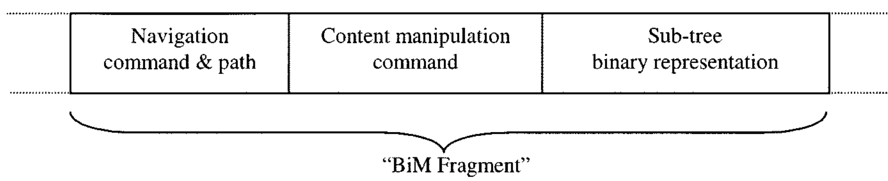

Fig. 3. Structure of a BiM fragment.

C. Binary Format

There are two main reasons for having a packet-based binary format for MPEG-7 descriptions. First, the textual format requires more storage or transmission resources than necessary from a theoretical point of view. Therefore, an efficient compression of the textual format is applied when converting it to binary. Second, the textual format is not very appropriate for streaming applications. For this type of application, a high level of flexibility is required with respect to the transmission order of the elements. Furthermore, easy random access should be provided without having to parse the complete bit stream. This requires a packet-based structure, for which headers are generally represented in binary format.

An MPEG-7 bitstream is composed of the so-called BiM fragments. Their structure is illustrated in Fig. 3. A BiM fragment is composed of three parts and conveys the same type of information as the textual access unit.

The first part is the "Navigation command and path" which signals to the terminal where to apply the MPEG-7 descriptions update. The addressing is specified by a navigation path, which defines the address of the tree node for which the decoder will receive the information contained in the BiM fragment. Several modes of addressing description elements are possible. For example, one can use a relative addressing with respect to the node the terminal is currently in or an absolute addressing with respect to the description root node.

The second part of a BiM fragment is the "Content manipulation command." It tells the decoder what kind of content manipulation it has to perform to combine the description that has already been reconstructed with the piece of information contained in the fragment. As in the case of access units, possible commands are "add," "update," and "delete."

The piece of MPEG-7 descriptions is contained in the third component of the BiM fragment. Its general binary format is composed of one global header and a set of consecutive and nested patterns. Patterns are composed of two main components, as illustrated in Fig. 4:

1) a pattern header, which is composed of a "key" specifying the format of the following data elements and of a "number of occurrence" specifying the number such data elements;
2) a pattern body, which is a set of n consecutive similar items, where n is the number of occurrences defined in the pattern header.

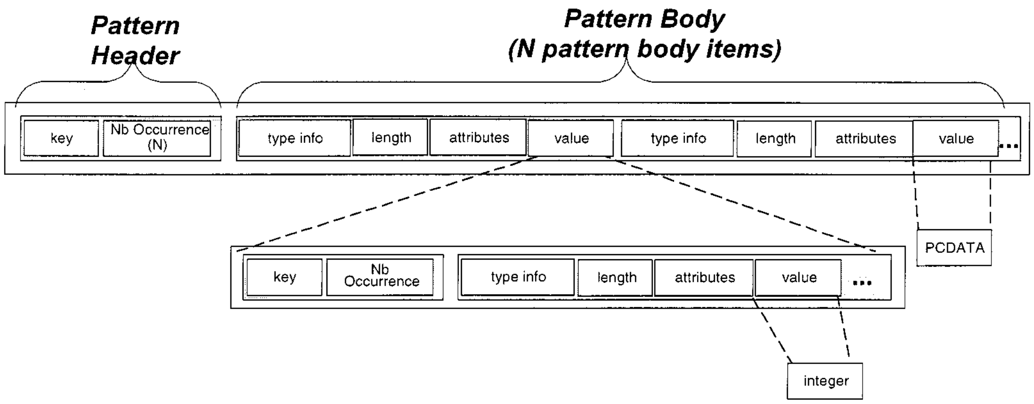

Fig. 4. Format of the sub-tree binary representation.

Each item of the body is composed of a "type info" encoding the XML Schema type of the element, a "length" describing the coding length in bits in order to allow a fast access of elements in the stream, "attributes" and a "value" encoding the description data. The value itself could be a pattern.

When the terminal knows the schema corresponding to the description, some components in the patterns are optional: for example, the "key" is not coded if the XML Schema element or attribute is mandatory. The "number of occurrence" is not coded if it is fixed in the schema. Attributes' data are not coded if the type cannot have attributes (e.g., XML Schema simpleType). The length is encoded only if the fast access feature is desired, etc.

D. Delivery of MPEG-7 Descriptions

The delivery of MPEG-7 description on particular systems is outside the scope of the MPEG-7 standard. Existing delivery tools may be used for this purpose. MPEG is developing specifications for the transport of MPEG-7 data on MPEG-2 Systems, as well as along MPEG-4 content. Transport of MPEG-7 description on other systems (e.g., analog delivery systems) may be similarly developed by appropriate organizations.

The transport of MPEG-7 data on MPEG-2 Systems shall be done according to a new amendment of the MPEG-2 Systems specification [5]. This amendment provides different means to carry metadata: a synchronous transport is provided by the carriage of metadata in Program Elementary Streams (PES) packets. Asynchronous transport without carousel delivery is provided by metadata sections, while the DSM-CC tools can provide carousel delivery mechanisms, with or without file structures.

The transport of MPEG-7 data along MPEG-4 content is done by considering MPEG-7 data as a specific kind of MPEG-4 elementary streams. The elementary stream identification for MPEG-7 data is already provided in the MPEG-4 specification, and the configuration information for such a stream is to be added to an on-going MPEG-4 amendment. From this amendment will derive the transport of MPEG-7 data on popular delivery mechanisms already available for MPEG-4, like IP networks and files using the MP4 file format.

IV. Conclusion

The tools described above contain the majority of the functionality of MPEG-7 Systems and allow the development of compelling MPEG-7 multimedia applications. These are provided by the current MPEG-7 standard, which will be international standard in September 2001.

They can be summarized as follows:

1) delivery, synchronization, and management of multimedia descriptions using a variety of transmission and storage protocols;
2) representation of dynamic multimedia descriptions through encapsulation of descriptions and schema structure in commands;
3) textual and binary representation of descriptions.

The technologies considered for standardization in MPEG-7 were not all at the same level of maturity. Therefore, as for MPEG-4, MPEG-7 has organized its specification in several phases, called versions. New versions complete the current standardized toolbox with new tools and new functionality. They do not replace the tools of the previous versions. This paper has described the MPEG-7 Systems Version 1 toolbox. Technology under consideration in MPEG-7 Systems Version 2 includes API's definition so that an application, possibly multi-user, can have access to the description and manipulates it according to its usage rights.

Acknowledgment

The MPEG-7 Systems specification reflects the results of teamwork within a worldwide project in which many people invested time and energy. The authors would like to thank them all, and hope that the paper accurately reflects the understanding of the group.

References

[1] *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 685–772, June 2001.
[2] *MPEG-7 Requirements Document*, ISO/IEC JTC1/SC29/WG11 N3933, Jan. 2001.
[3] *Text of ISO/IEC CD 15938-1 Information Technology—Multimedia Content Description Interface—Part 1 Systems*, ISO/IEC JTC1/SC29/WG11 N3701, Oct. 2000.
[4] *Text of ISO/IEC CD 15938-2 Information Technology—Multimedia Content Description Interface—Part 2 Description Definition Language*, ISO/IEC JTC1/SC29/WG11 N3702, Oct. 2000.
[5] *Text of ISO/IEC 13 818-1:2000/PDAM1—Transport of Metadata*, ISO/IEC JTC1/SC29/WG11 N3845, Jan. 2001.
[6] J. Hunter *et al.*, "MPEG-7 Description Definition Language," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, pp. 765–772, June 2001.