FROM DESCRIPTIONS TO CONTENT: INVERTING THE SENSE (AUTHORING744)

José M. Martínez and Francisco Morán

Grupo de Tratamiento de Imágenes, E.T.S. Ing. Telecomunicación
Universidad Politécnica de Madrid, E-28040 Madrid, Spain

Tel: +34.913.367.353; Fax: +34.913.367.350
E-mail: {jms,fmb}@gti.ssr.upm.es

ABSTRACT

This paper presents a starting point for trying to use MPEG-7 to synthesise content instead of analysing it, that is, for using content descriptions to generate (synthetic) content, instead of creating descriptions from existing content. The output synthesised content will use MPEG-4 (XMT) as the representation format.

Keywords: authoring tool, MPEG-7 content description, MPEG-4 content representation.

1. INTRODUCTION

The new MPEG standards MPEG-4 [1] and MPEG-7 [2] are aimed at providing new ways for authoring, transporting, accessing, retrieving and interacting with multimedia content. MPEG-4 is the first standard supporting interactive and complex multimedia presentations. MPEG-7 is targeted at standardising the results of the analysis of existing content.

This paper introduces the first ideas for trying to use MPEG-7 to synthesise content, instead of analysing it, that is, for going from content descriptions to (synthetic) content, instead of the other way around. The synthesised content will use MPEG-4 (XMT [3]) as the representation format.

2. MPEG-4 AND MPEG-7 FUNCTIONALITIES

MPEG-4 [1] is a standard for multimedia applications that supports the creation of rich, reusable, interactive, and transport-transparent multimedia content. MPEG-4 reached the ISO standard status in 1998 for version 1 and in 1999 for version 2; and is currently working in further extensions (versions 3 and 4). MPEG-4 provides authors with normalised tools for creating content with far greater reusability; it provides network service providers with transparent information that can be interpreted and translated into the appropriate native signaling messages of each network with the help of relevant standards bodies; and it provides users with higher levels of interaction with content, within the limits set by the author. MPEG-4 XMT [3] provides a textual representation of MPEG-4 content, easing the presentation using SMIL, VRML and MPEG-4 players.

MPEG-7 [2], to become an ISO standard in the Autumn of 2001, aims at normalising AV content description in order to facilitate information search (by query and/or by filtering). MPEG-1, -2, and -4 were designed to represent the information itself, while MPEG-7 is meant to represent information about the information. Another way of looking at it is that MPEG-1, -2, and -4 make content available, whereas MPEG-7 allows to find some needed content.

MPEG-7 compliant descriptions are associated with the content they describe, to allow fast and efficient searching for and filtering of material that is of interest to the user. The standard does not comprise the (automatic) extraction of descriptions/features, nor does it specify the search engine, filter agent, or any other application that can make use of the descriptions. In principle, any type of AV material may be retrieved by means of any type of query material. This means, for example, that visual material may be queried using any kind of material: visual, music, speech, etc. It is the responsibility of the search engine and filter agent to match the query data to the MPEG-7 AV description. Nevertheless, search and filter applications are not the only ones that may benefit of MPEG-7.

MPEG-7 standardises the description of various types of multimedia content. These descriptions are based on different features of the content, be it catalogue attributes (e.g., title, creator, rights), or semantic (e.g., the "who/what/when/where" information about objects and events) or structural (e.g., the colour histogram of an image or the timbre of a recorded instrument) characteristics. It also covers the collections of objects and tools for navigation, access and user interaction.

3. THE VISION: AUTHORING CONTENT THROUGH MPEG-7 DESCRIPTIONS

Complex authoring tools are required in order to create the powerful reusable and interactive content allowed by MPEG-4. On the other hand, MPEG-7 aims at describing content in a more natural manner. Currently,

developments of MPEG-7 applications extract the characteristic features of some existing content in order to generate a standardised description of it. This is the straightforward way of using MPEG-7 — or maybe just the first one envisioned?

Although it is clear to the popular wisdom that a picture is 1000 words worth, MPEG-7 may help compression too. MPEG-7 provides the tools for having compact, efficient and normalised descriptions of content, and it may (should?) be possible to create the content from descriptions. In principle the description will need less bits for its representation, so it would be possible to transmit the description and then reconstruct the content in the (presentation) terminal.

Creating AV interactive content, which means now using complex (but powerful) authoring tools, could mean writing descriptions from which to generate content automatically. The content can be generated either at the server side or at the client side, allowing to choose between compressed storage and/or transmission and computing resources for transcoding the description to content.

Of course, this vision may yield only drafts of the AV content that will bring to reality the author's idea, but it will ease the quick creation of first prototypes, e.g., advertisement images, soundtracks, jingles, story-boards for pilot series, publicity/music video-clips, etc. Nevertheless, in some cases, it should be possible to create final content.

4. THE AUTHORING744 RESEARCH INITIATIVE

The main objective of the Authoring744 research initiative is the development of an operational prototype allowing the creation of AV content through writing scripts/descriptions in a user-friendly GUI. By "script writing" we mean filling forms in and choosing some media description examples (e.g., the key of a melody, the texture of an object, the dominant colour of a region).

The output of the *ScriptWriter Tool* will be MPEG-7 descriptions of the content. A *724Transcoder Tool* will take this output and create the actual MPEG-4 XMT files. The advantages of using MPEG-4 XMT [3] are that it uses XML-Schema [4] and that it is designed to be easily transcoded, in turn, into MPEG-4 BIFS, VRML, or SMIL coded multimedia content. This will allow displaying the content in various terminals and within different applications (VRML and SMIL web plug-ins are available).

The major milestones, and some open questions, are:

1. MPEG-7 and MPEG-4 XMT harmonisation: Definition of the MPEG-7 descriptions that can be matched to MPEG-4 XMT elements, and identification of description tools which are missing in MPEG-7 and required for creating a complete MPEG-4 XMT content representation.
2. Development of the *ScriptWriter Tool*: Is it affordable to have only one generic authoring application, or is it better to focus on a few specific ones targeted at different kinds of output content (e.g., images, story-boards, sound-tracks)? We think the second choice is the more realistic one, at least for a first approach.
3. Development of the *724Transcoder Tool*: This is where the actual "724" (MPEG-7 to MPEG-4) step is performed.
4. Development/adaptation of new/existing MPEG-4 XMT transcoders for producing MPEG-4 BIFS, VRML, and/or SMIL coded content.
5. Selection of players (MPEG-4, SMIL, VRML), integration of applications, tests and evaluation of results.

In the following sections we will describe the current status in each of the above working areas.

4.1. MPEG-7 and MPEG-4 matching

4.1.1. Current MPEG-7 tools

In order to create content from description, not all of the MPEG-7 description tools are to be taken into account, at least in a first phase. MPEG-7 description tools currently cover six viewpoints (see Figure 1 [5]):

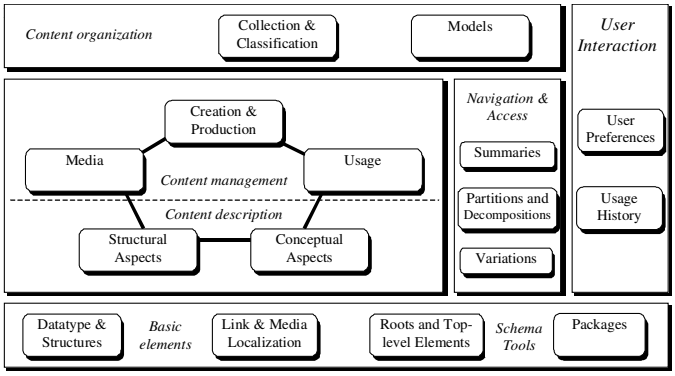

Figure 1. Overview of MPEG-7 description tools

1. *Basic elements and schema* tools provide support for building the "real" descriptions, so no implicit matching is required.
2. *Content management* tools cover the description of the media, the creation and classification of content, and the usage information. In order to create content, most of this description tools have no direct use. However, some of them could be used to support IPMP functionalities; and the identification of actors, presenters, etc., might help to synthesise characters in the content (or to create characters that would resemble the desired one).

3. *Content description* tools cover the description of the content, and these are the most targeted ones to our objective: the description of the content to be generated. We will analyse these in more detail in section 4.1.2.
4. *Content organisation* tools describe collections of objects, and although they may have future use (creation of similar contents), currently they are not under consideration for Authoring744.
5. *Navigation and access* tools allow to create alternative presentations of the content in terms of resolution views (in space and frequency), summaries, and transcoded variations (in resolution or media translation). These tools, although not directly implied in the generation of content, may be useful in providing hints for generating the alternative presentations of the content depending on the terminal, user and network preferences.
6. The *User interaction* tools in MPEG-7 cover currently only the description of user preferences and usage history, so they do not provide tools allowing the creation of content with interaction options (as one may expect from the viewpoint name).

In short, from the current MPEG-7 description tools, the only relevant to the Authoring744 objective are the *Content description* tools, although some description tools from *Content management* and *Navigation and access* may provide added-value functionalities.

4.1.2. MPEG-7 Content description tools

The current main description tools in MPEG-7 for describing content are grouped in Structural and Conceptual aspects description tools.

Structural description tools describe content in terms of segments: temporal, spatial and spatio-temporal ones. Each segment entity is described, besides by its temporal and/or saptial boundaries, by management and annotation description tools and, what is more relevant for Authoring744, by low-level audio-visual description tools (colour, shape, texture, motion, faces, melody, sound effects, timbre, ...). So, it may be possible upon some extended to synthesise content from Structural descriptions. The limit being that these descriptions are low-level, that is, they allow the description in terms of syntactic structure and perceptual features. Nevertheless, it should be possible to write a description in terms of regions (both spatial and temporal) with associated low-level features (even using media examples for writing the description, that is, choosing among colours, textures, ... from a given set or even allowing editing the media examples). In this case, the *ScriptWriter Tool* seems to be oriented to be a Graphics and Sounds Editor, using MPEG-7 as intermediate format.

At this point, we have a standardised description driven Multimedia Objects Editor, that will allow to store and transmit the content in a compressed form (as said before, the description will be smaller than the content itself).

Conceptual description tools aims to describe content in terms of events, (semantic) objects, places and times, and their relationships. Also, the Conceptual and Structural entities can be associated, providing semantic meaning to the different segments described via structural and perceptible features. Using these description tools that allow to describe the content using high-level abstractions, the *ScriptWriter Tool* becomes more than a Multimedia Object Editor, as semantic meaning and behaviours and actions may be described. The current challenge is then how to translate this for synthesising content, that is, the required mapping procedure to MPEG-4 XMT.

4.1.3. Current MPEG-7 tools are not enough

On the other hand, we have already identified the need for additional description tools. For instance, as the SNHC (Synthetic/Natural Hybrid Coding) subgroup of MPEG has only produced MPEG-4 tools up to date, MPEG-7 does not support yet the description of synthetic content (there is only a 3D Shape Visual Descriptor which allows the intrinsic shape description of 3D mesh models, but targeted to search and browsing of 3D model databases [6]). This is especially unfortunate from the Authoring744 viewpoint, because "synthesising synthetic content" should be of course easier than "synthesising natural content". It should be far more straightforward to have a synthetic talking head resembling Schwarzenegger's recite Shakespeare's "to be or not to be" than to have Scwharzenegger himself play Hamlet...

The FBA (Face and Body Animation) activities of the SNHC subgroup have already yielded some tools for the description of high-level face and body expressions that have been adopted in MPEG-4 versions 1 and 2. Mechanisms have been designed, for instance, to deform the 3D mesh representing a face to make it express anger, happiness, sorrow, etc., by grouping FAPs (Face Animation Parameters) in "visemes", in order to improve coding efficiency. FAPs can already be regarded as middle-level control knobs to deform the detailed 3D mesh; visemes would then be meta-FAPs.

Furthermore, the SNHC subgroup is focussing now on the AFX (Animation Framework eXtension). And although AFX activities are also targeted mainly at MPEG-4 (of which new versions featuring new functionalities keep appearing) in the short term, its middle (long?) term objective is to produce high-level, semantic descriptions of animations [6].

This is very much what we would need as an output of Authoring744's *ScriptWriter Tool*, if we want to create content including expressive characters, but

MPEG-7 currently lacks such high-level tools for describing face features and gestures.

4.2. MPEG-7 ScriptWriter

We are in the design phase of the GUI of the MPEG-7 based *ScriptWriter*. We are working in the metaphors that shall be used for the selected MPEG-7 description tools and the identified additional description tools. The development of the application will start after the final selection (in base of the target content) of MPEG-7 and additional description tools.

4.3. 724Transcoder

The development of the *724Transcoder* (from MPEG-7 descriptions to MPEG-4 XMT content) will also start after the selection of description tools.

The *724Transcoder* will be designed and implemented in order to allow both its use at the server side (allowing to store and/or transmit MPEG-4 XMT content) and the terminal side (allowing to reduce the size of storage and bit rate by storing and transmitting the description, that will then be transcoded just before presentation).

4.4. Playing XMT content

In order to play XMT content, the files are translated to one of the following content representation standards (see Figure 2 [3]): SMIL, VRML, and MPEG-4. The XMT content (representation) can be parsed and played directly by a W3C SMIL player, preprocessed to Web3D X3D and played back by a VRML player, or compiled to an MPEG-4 representation such as mp4, which can then be played by an MPEG-4 player.

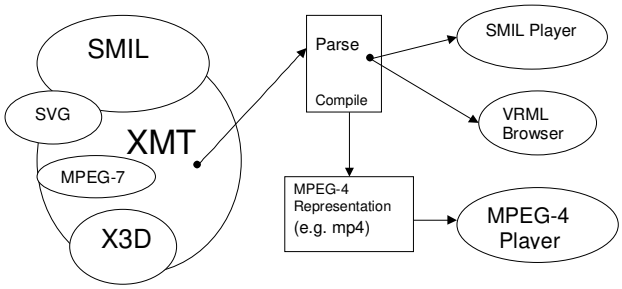

Figure 2. Playing XMT content

4.5. Integration and Tests

At the current stage there is nothing to integrate, neither to test.

5. CONCLUSIONS

This paper has presented the vision and current status of the Authoring744 research initiative, aiming to create content from high-level descriptions.

MPEG-4 and MPEG-7 have been selected, respectively, for content and description representation. At this preliminary stage of the research, we can conclude that it is possible to use this authoring methodology (writing descriptions to be transcoded to content) in order to generate content as in a Multimedia Object Editor. We have already started to explore how to translate the Conceptual information to be provided via MPEG-7 descriptions to MPEG-4 content in order to create content using more abstract and generic components and concepts in the *writing* process. We have identified the need of further description tools (description of synthetic objects) in order to provide new functionalities allowing the authoring of more generic and powerful content.

REFERENCES

[1] R. Koenen (ed.), "MPEG-4 Overview", MPEG doc. N3931, January 2001.

[2] J.M. Martínez (ed.), "Overview of the MPEG-7 Standard", MPEG doc. N3752, October 2000.

[3] M. Kim, M. Bourges-Sévenier, R. Rafey, S. Wood (eds.), "ISO/IEC 14496-1:2000/PDAM2", MPEG doc. N3725, October 2000.

[4] D.C. Fallside (ed.), "XML Schema Part 0: Premiere", W3C Candidate Recommendation, October 2000.

[5] P. van Beek, A.B. Benítez, J. Heuer, J.M. Martínez, P. Salembier, J.R. Smith, T. Walker (eds.), "CD 15938-5 Information Technology: Multimedia Content Description Interface: Part 5: Multimedia Description Schemes", MPEG doc. N3705, October 2000.

[6] A. Yamada, M. Pickering, S. Jeannin, L. Cieplinski, J.R. Ohm, M. Kim (eds.), "MPEG-7 Visual Part eXperimentation Model (version 9)", MPEG doc. N3914, January 2001.

[7] M. Bourges-Sévenier, H. Grahn, B. Miollan, F. Morán, I.A. Salomie, M. Steliaros, A.E. Walsh (eds.), "MPEG-4 AFX Verification Model (version 2.0)", MPEG doc. N3922, January 2001.