MULTIMEDIA ADAPTATION BY TRANSMODING IN MPEG-21

Mariam Kimiaei Asadi and Jean-Claude Dufourd

ENST Paris Dept. Comelec, 46 Rue Barrault, 75634 Paris.
{kimiaei, dufourd}@enst.fr

ABSTRACT

Transmoding is an important part in the content adaptation process of a Universal Multimedia Access system. In this paper we introduce a description tool for online transmoding in MPEG-21. We then give some use case scenarios for adaptation by transmoding. We also demonstrate the efficiency and the necessity of the transmoding description tool through some use cases and the corresponding applications. Our implementation of these applications is based on our approach of MPEG-21 reference software to which the transmoding description tool has been added. Also we will describe the integration of this implementation to an IST European project. Finally we end up by giving a perspective for future integration of transmoding-type description tools in MPEG-21.

Keywords: modality, multimedia, transmoding, transcoding, modality conversion, universal access.

1. INTRODUCTION

Today, to deliver and consume multimedia content, one needs to have a multimedia adaptation infrastructure. Without such an infrastructure, content creators and service providers face several problems in providing their consumers with multimedia content. One of the aims of MPEG-21 standard is to normalize a framework for multimedia adaptation. The MPEG-21 multimedia adaptation infrastructure is built upon several elements. Some of these elements are the constraints and characteristics of the environment in which the concerned multimedia content is going to be used. This concept is called "Usage Environment" characteristics and contains different constraints such as user terminal capabilities, network characteristics, user and author preferences as well as constraints imposed by any other existing intermediate peer in a multimedia content delivery chain.

The adaptation should be done with respect to all of these constraints. The adaptation is then either transcoding, transmoding, any other parameter change of the original content or a combination of all these kinds of adaptation.

Although adaptation by transmoding may be the solution for many frequent multimedia content adaptation use cases, so far, the focus of most researches on content adaptation has been limited to adaptation by transcoding and its needed support. As a result, support for transmoding-type adaptations has not been sufficiently investigated.

In this paper we define a description tool for online transmoding in MPEG-21 and we show its efficiency and necessity through several use case applications. In the next section we first give a short introduction on MPEG-21. In section III We give an exact definition for "modality" and "transmoding" notions and based on that we introduce the transmoding table. In section IV we discuss the need and necessity of transmoding support in MPEG-21 by bringing several use case scenarios. In section V we present our transmoding description tool. In section VI, we introduce our implementation of several transmoding applications, which is developed using our approach of MPEG-21 reference software, we also describe the integration of this implementation into an IST European project on multimedia adaptation, named ISIS (Intelligent Scalability for Interoperable Services), which funds part of our research on MPEG-21. In section VII we give our conclusion on this work and finally in section VIII we talk about the perspective and continuation of this work in MPEG-21.

2. MPEG-21

MPEG-21 is an ISO standard from MPEG family who identifies and defines the key elements needed to support a multimedia delivery chain, the relationships between and the operations supported by them. Within the parts of MPEG-21, MPEG elaborates these elements by defining their syntax and semantics. MPEG-21 also addresses the necessary framework functionality, such as the protocols associated with the interfaces, and mechanisms to provide a repository, composition, conformance, etc. A basic notion of MPEG-21 is "Digital Item" (DI). A "Digital Item" is a multimedia content and as described in MPEG-21 a "Digital Item" is the digital representation of "a work", and as such, it is the thing that is acted upon

(managed, described, exchanged, collected, etc.). Some main parts of MPEG-21 are Digital Item Declaration (DID) [2], Digital Item Adaptation (DIA) [3][4] and Right Expression Language (REL) [5].

3. TRANSMODING AND MODALITIES

To define transmoding, we first need to come up with an exact definition for "modality" and a list of modalities.

Modality has at least two meanings: on the perceptual level, modalities are tied to the five human senses, thus there is only "one" visual modality; on the structural level, there may be many modalities within one (perceptual) modality. For example, visual modalities include bitmap video, bitmap image, and two flavors of vector images: graphics2D and graphics3D. Within MPEG-21, the right meaning is the structural one, i.e. modality from the point of view of adaptation, since it dictates the types of algorithms, which can be applied to the resources.

We have proposed a hierarchical schema for the Modality Classification Schema in a Core Experiment on transmoding. In this classification schema, we have five principal modalities and some sub-modalities for some of them: Video, Audio (Audio2D, Audio3D, Speech), Image, Graphics (Graphics2D, Graphics3D) and Text.

Therefore, we define "Transmoding" as a kind of a digital resource adaptation that changes the modality of the original resource. In this paper transmoding refers to online modality conversion of a multimedia resource and not offline modality change based on the substitution of a resource by existing alternate versions of it in other modalities. In figure 1, all (online) transmodings for the above-defined modalities are illustrated. The white boxes are the transmodings that seem to make sense, be useful and have general parameters which are not algorithm-dependent. The grey boxes are the other ones. There are some transmodings, which have not been investigated in this work and need to be investigated by related modality experts; the related boxes contain a question mark.

In this paper, Graphics modality refers to vector graphics and Image modality refers to bitmap image.

4. SOME USE CASE SCENARIOS

A use case scenario for Video-to-Image and Video-to-Ggraphics2D transmoding is the case of an original resource of Video modality for which the provider has not provided the image or slideshow alternatives but has given his intentions concerning the transmoding to Image and Slideshow (Graphics2D) through the transmoding descriptor. For the Video-to-Slideshow conversion, the provider has determined which frames to use as slides, the slides transitions and durations.

OUT \ IN		Audio			Image	Video	Graphics		Text
		Speech	Audio2D	Audio3D			Graphics2D	Graphics3D	
Audio	Speech		?	?					?
	Audio2D	?		?					
	Audio3D	?	?						
Image		*				√	√	√	√
Video							√	√	√
Graphics	Graphics2D					√		√	√
	Graphics3D	?							?
Text		?	**	**	**	**	**	**	

*This transmoding could be done in different ways, one way would be Speech to Text + Text to Image, the other way would be Speech to sign language.
**This transmoding does not make sense as an online conversion.

Figure 1: Transmoding table.

A concrete example of this scenario is as follows, the original video resource is a movie trailer which is designed for ADSL but is not suitable for cellular networks. In the corresponding REL expression, Video-to-Slideshow transmoding is authorized. The optimum adaptation decision, which is taken on the basis of all related constraints, is transmoding to Slideshow. Once this decision is taken, to perform the action of transmoding, the needed metadata is fetched from the DIA descriptor. For example, the display resolution and the format of the image are fetched from TerminalCapabilities descriptor. Any constraint on modality and modality-related priorities are fetched either from the ModalityConversionPreference descriptor or PresentationPriorityPreferences descriptor. The available bandwidth which is needed to determine the compression level of the image sequence is fetched from the NetworkCharacteristics descriptor. And finally the list of which frames to use and the durations, is fetched from the transmoding descriptor.

A use case scenario for Text-to-Image transmoding could be as follows. The resource to be adapted is a text resource: for example, a message in the Persian language (or with any special font, family and style). The device is a simple one which does not have the Persian font support, but supports image content. In this case, if authorized in REL expression, a solution is to transmode the text to an image. The transmoding process requires font characteristics (to render the text into an image), color information, etc. The same use case may be given for a Text-to-Video transmoding, when the text does not

fit into the screen of the terminal (whose size is extracted from TerminalCapabilities descriptor), and thus needs to be scrolled. This transmoding process additionally requires the author preferences on the type of scrolling.

5. TRANSMODING DESCRIPTION TOOL

With the purpose of producing the metadata on the adaptation of a resource by transmoding, we present in this paper, a description tool for transmoding. We are interested in the point of view of the provider. Nevertheless, through this descriptor, any peer in the adaptation chain can express his preferences on the parameters of a specific conversion. The provider, based on his knowledge on the resources, may need to provide some transmoding hints to facilitate, guide or enable the adaptation; this will be done through the proposed descriptors. The concept of transmoding descriptor within DIA is the same as network or terminal capabilities descriptors, which provide the needed metadata for adaptation process based on the given constraints.

The transmoding hints include the descriptions of the most general transformation parameters i.e. the descriptors are based on no particular underlying algorithm. A transmoding descriptor is not supposed to contain any transmoding-related right or permission statement. The corresponding right and authorization statements will be given through MPEG-21 REL [5].

Figure 2 presents a concrete example of transmoding descriptor for Text-to-Image conversion.

6. IMPLEMENTATION

In the implementation of the transmoding description tool, first we developed parsing software within the framework of the MPEG-21 DIA reference software, that performs the parsing job of the Usage Environment Descriptor (UED) and transmoding descriptor. In the second step, we developed a decision making software and added it, together with a set of transmoding tools, to the parser software. The whole software was contributed to MPEG in a Core Experiment. This implementation was also integrated to an IST European project on multimedia adaptation, called ISIS [6].

The ISIS transmoding module is shown in figure 3 and is composed of two main blocks: the Decision Maker and Resource Transmoder. Its inputs are User's selection of content from the Streaming Server (DID), the description of the content from a Content Description Database and the user's constraints description from a Temporary Cache Database.

The Decision Maker is the decision making algorithm which determines the optimal transmoding form for the resources with certain constraints. The Decision Maker

```
<Transmoding>
   <TransmodingMethodParameters
    xsi:type="TextVisualizationParametersType"
    quality="1.0">
      <To
href="urn:mpeg:mpeg21:cs:ModalityCS:2003:2">
         <mpeg7:Name>Image</mpeg7:Name>
      </To>
      <FontParameters fontSize="20">
         <FontStyle
href="urn:mpeg:mpeg21:cs:FontStyleCS:2003:2">
            <mpeg7:Name>BOLD</mpeg7:Name>
         </FontStyle>
         <FontFamily
href="urn:mpeg:mpeg21:cs:FontFamilyCS:2003:2">
             <mpeg7:Name>ARIAL</mpeg7:Name>
         </FontFamily>
      </FontParameters>
      <TextColorInformation
            foregroundColor="0.2 0.5 0.6
            backgroundColor="0 0 0"/>
      <Resolution horizontal="300"
                  vertical="400"/>
   </TransmodingMethodParameters>
</Transmoding>
```

Figure 2: a concrete example of transmoding descriptor for text-to-image conversion

has been entirely developed in java language, and comprises of three main modules:

- A *DID Parser* (origin: Ghent University)[7].
- The *DIA Parser* (UED and transmoding descriptor Parser) which has been developed based on a DOM-like approach.
- The *Transmoding Decision Making Logic* which is the decision making engine of the Decision maker. It takes as input the constraints stored into an internal structure and the resource references and is based on a rule-based approach. This process comprises of two main steps. In the first step a small number of constraints are considered, which are the ones needed for making any transmoding decision, for example modality supports in terminal capabilities descriptors. These constraints will be used to filter out all transmoding options that are not supported by the current terminal and network. In the second step, from the remaining options that are technically feasible, an optimal decision is taken using the remaining constraints.

The Resource Transmoder is an implementation of a set of transmodings: Text to Image, Graphics2D to Video, Graphics2D to Image, and Video to Graphics2D which have been implemented in Java or C. The inputs of the Resource Transmoder are the transmoding decision, the reference of the resource to be transmoded and the required parameters to perform the related transmoding.

We have done some measurements for three different transmoding tools. The OS is Windows™ 2000, the processor is an AMD Athlon™ XP 2000. The target display dimension is assumed to be 176x144. The average execution times are: 2800 ms for Graphics (BIFS

213Kb, 15s, 12fps) to Video (MPEG-4 Video), 420 ms for Text to Image (mostly PNG encoding) and 15 ms for Graphics to Image conversion. Please note that our tools are not optimized, and that these numbers are only indicative.

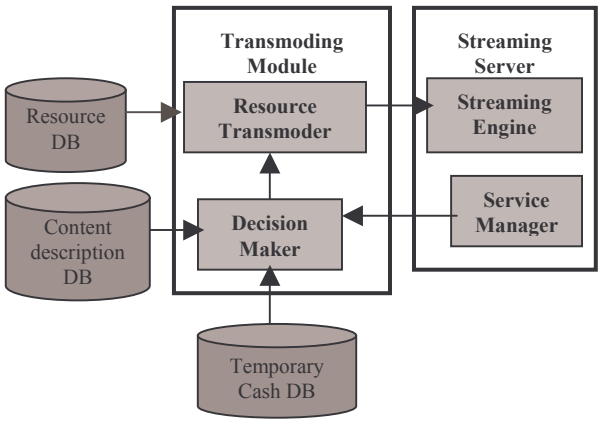

Figure 3. ISIS Transmoding module architecture

7. CONCLUSION

In this work, the transmoding description tool was shown to be useful and in some cases necessary in order to construct the support of transmoding resource adaptation in MPEG-21. This description tool, which is based on the provider's knowledge of the resources, is basically suitable for the provider to express his preferences and hints on different transmoding parameters. Through our developed software, which is based on our approach of MPEG21 reference software, the presence of the transmoding description tool in MPEG-21 DIA was demonstrated to be crucial in facilitating and guiding the adaptation process for the conversion of the resources modalities.

8. PERSPECTIVES

Recently, in MPEG-21, as the continuation of our work on the transmoding description tool, the need for concrete descriptions of conversions (as opposed to just constraint description) was recognized. The transmoding description tool was deemed to be an example of such concrete descriptions and it was proposed to extend it to any other kind of resource adaptation (transcoding or any other parameter changes).

Our future work will therefore concern this need and will include the research on a complete conversion description tool and its implementation. A sample DID instance showing the concept of conversion description tools is presented in Figure 4.

```
<DIDL>
 <Item>
  <Component>
     <Descriptor>
       <Statement mimeType="text/xml">
         <dia:ConversionInformation>
             <dia:Conversion>
                <dia:Transcoding>

<!—here go transcoding descriptors expressing
the hints for Transcoding of this video to MPEG-
4 Video Advanced Simple Profile for example: -->

                  <dia:To>
            MPEG-4 Video Advanced Simple Profile
                  </dia:To>
                  <dia:With>
                      ...Some parameters...
                  </dia:With>
                </diaTranscoding>
                <dia:Transmoding>

<!—Here go transmoding descriptors expressing
hints for conversion to image for example:-->

                   <dia:To>Image</dia:To>
                   <dia:With>
                      ...Some parameters...
                   </dia:With>
                </dia:Transmoding>
             </dia:Conversion>
         </dia:ConversionInformation>
       </Statement>
     </Descriptor>
     <Resource mimeType="video/mp4"
               ref="myInitialVideo.mp4" />
   </Component>
 </Item>
</DIDL>
```

Figure 4: a DID sample with conversion description tools

9. REFERENCES

[1] R. Mohan, J.R. Smith and C.S. Li, "Adapting Multimedia Internet Content for Universal Access", *IEEE Trans. Multimedia*, Vol. 1, No. 1, pp. 104-114, Mar 1999.

[2] MPEG MDS Group, MPEG-21 Digital Item Declaration FDIS, ISO/IEC JTC 1/SC29/WG11/N4813, Fairfax, VA, USA, May 2002.

[3] MPEG MDS Group, MPEG-21 Digital Item Adaptation FCD, ISO/IEC JTC 1/SC29/WG11/N5845, Trondheim, Norway, July 2003.

[4] MPEG MDS Group, MPEG-21 Digital Item Adaptation AM 6.0, ISO/IEC JTC 1/SC29/WG11/N5846, Trondheim, Norway, July 2003.

[5] MPEG MDS Group, MPEG-21 Rights Expression Language FCD, ISO/IEC JTC 1/SC29/WG11/N5839, Trondheim, Norway, July 2003.

[6] http://isis.rd.francetelecom.com

[7] http://www.titr.uow.edu.au/cgi-bin/mpeg-ref-sw.pl