INTRODUCING CAIN: A METADATA-DRIVEN CONTENT ADAPTATION MANAGER INTEGRATING HETEROGENEOUS CONTENT ADAPTATION TOOLS *

José M. Martínez, Victor Valdés, Jesús Bescós, Luis Herranz

Grupo de Tratamiento de Imágenes
Escuela Politécnica Superior, Universidad Autónoma de Madrid, SPAIN
E-mail: {JoseM.Martinez, Victor.Valdes, J.Bescos, Luis.Herranz}@uam.es

ABSTRACT

This paper introduces a content adaptation manager targeted to the integration of different metadata-driven content adaptation approaches. The main objective is to provide a unified content adaptation module with media, media description and context description inputs that will select the most appropriate adaptation parameters and afterwards select to most suited content adaptation tool. The MPEG descriptions standards (MPEG-7 and MPEG-21) have been selected as metadata representation formats. The proposed Content Adaptation Module would be used for service prototyping (and deployment) and as a framework for benchmarking of different content adaptation tools. Current status of the developments within the IST aceMedia project is described.

1 INTRODUCTION

Content Adaptation is the main objective of a set of technologies that can be grouped under the umbrella of the Universal Multimedia Access (UMA) concept[1]. This refers to the capability of accessing to rich multimedia content through any client terminal and network.

The development of both new access networks providing multimedia capabilities, and a wide and growing range of terminals makes the adaptation of content an important issue in future multimedia services. Besides, in order to enhance the user's experience[2], not only terminal and network parameters, but also user personalization and environmental conditions should be taken into account when adapting.

This paper presents an overview of the functional design, architecture, and current implementation status of a content adaptation manager targeted to the transparent integration of different but complementary content adaptation approaches under a single module: the content adaptation manager named CAIN module (Content Adaptation Integrator).

The CAIN will be mainly invoked when the user wants to adapt content for presentation. Additionally, as the user is browsing the set of results, the application module providing this browsing functionality may call the CAIN to require adaptation of the corresponding browsing oriented media (e.g., keyframe, trailer, etc.).

The CAIN is being developed within the aceMedia Project[3]. The project is built around the concept of an ACE, an autonomous content entity, which is able to adapt itself depending on the varying target conditions described via user, platform, and network profiles, and on automatically extracted knowledge of its own content, described via a set of metadata. The aceMedia system aims to research, develop and integrate the wide range of technological modules required to support the establishment of the ace concept, and the CAM is one of the main modules involved in content adaptation.

Adaptation objectives focus on transcoding based adaptation, scalable content based adaptation (*"…scalable coding and transcoding … are technologies that meet different needs in a given application space and it is likely that they will coexist."* [4]), semantic driven adaptation[5] and transmoding approaches.

* Work partially supported by the European Commission under the 6th Framework Program, project FP6-001765: Integrating Knowledge, Semantics and Content for User-centred Intelligent Media Services (aceMedia). This work is also supported by the Ministerio de Ciencia y Tecnología of the Spanish Government under project TIC2002-03692 (DYMAS).

The CAIN aims to adapt content in the most efficient way, by means of integrating different content adaptation approaches, and establishing a compromise between the computational cost, the quality of the final adapted media, and the constraints imposed by the media formats. More specifically, this module acts as a processing node in charge of:

1. Selecting the appropriate media parameters and modality according to: media related user preferences, terminal characteristics, network capabilities, media information, media variations and summaries, and semantic annotations (e.g., regions of interest).
2. Selecting the appropriate available Content Adaptation Tool –CAT-.
3. Configuring and managing the selected CAT in order to create adapted content to be delivered to the appropriate content repository or streaming service.

Key technologies within the CAIN include MPEG-7 (mainly the Multimedia Description Schemes –MDS- specification[6]), MPEG-21 (mainly the Digital Item Adaptation –DIA- specification[7]), and available media transcoding and transmoding tools and scalable coding tools, that conform to the "delegate" content adaptation modules.

2 ARCHITECTURE OF THE CAIN

Figure 1 shows an overview of the CAIN architecture.

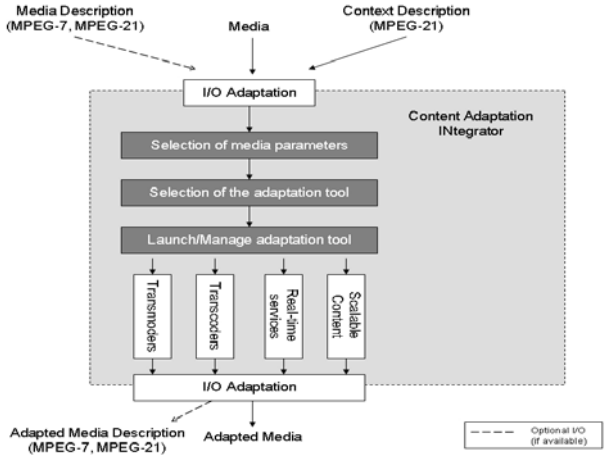

Figure 1: CAIN architecture

In response to an external invocation (via the corresponding API) the engine requests and receives content, MPEG-7 compliant content descriptions and MPEG-21 compliant bit streams descriptions. In parallel, the engine receives a context description compiled by the aceMedia User Profile (see section 3.2) which contains, at least, media related user preferences, terminal capabilities and network characteristics. The profile is then parsed in order to confront the selection of media parameters that would fulfil the adaptation and to decide which of the available CATs should be used. CATs are then launched and managed in order to produce adapted content and metadata which will be transferred to the delivery services.

The content adaptation modules correspond to different adaptation possibilities or categories (there can be several modules of the same category):

- The 'Transcoders' module represents decode-encode model transcoding.
- The 'Scalable Content' module will be in charge of truncation (limited expansion –e.g. interpolation- will be studied for further versions) of scalable content. It will use a format agnostic approach, accessing the bitstreams via a generic bitstream transcoding module (MPEG-21 DIA BSD/gBSD).
- The 'Real-time services' module aims to extract, via the use of real time analysis techniques, semantic features from compressed content in order to generate a semantic-driven adapted version.
- The 'Transmoding' module is intended to provide different kinds of transmoding (e.g., "simple" audiovisual to audio, video to slide-show transmoding, media to text (or voice) transmoding)

In order to integrate the CAIN into the aceMedia system architecture (or any other system), the CAIN provides some external interfaces (see **Figure 2**) to interact with other application modules.

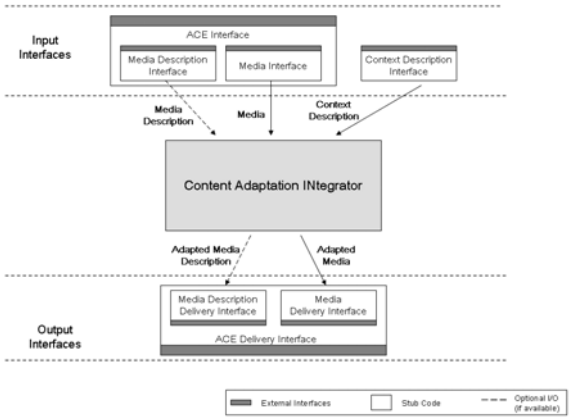

Figure 2: External interfaces to the CAIN

3 CURRENT STATUS

This section presents the current status of the implementation of the CAIN. Section 3.1 discuss shortly the media and formats under consideration. Section 3.2 gives some details of the description parsing and of the parameters and CAT selection process, whilst section 3.3 presents the different CATs currently integrated.

3.1 Media domains and formats

Media resources under consideration (domains and formats) depend on aceMedia selected scenarios. Regarding images, JPEG2000 has been selected, in order to have scalability features. In the case of video, aceMedia will consider for the first phase MPEG video, along with preliminary developments on scalable video[8]. Table 1 depicts a first mapping between Media types currently supported by CAIN and the category of CAT selected to manage their content adaptation. The CATs that are currently integrated are described in section 3.3.

MEDIA TYPE	CAT category
MPEG-1/2/4 SP video, MPEG-1 audio.	***Transcoding***
JPEG2000, SVC (aceMedia Scalable video).	***Scalable Content***
MPEG-1/2.	***Semantic driven***
MPEG-1/2/4 SP video.	***Transmoding***

Table 1: Relationship between Media types and CAT categories

3.2 Descriptions Processing

The content descriptions are based on MPEG-7 MDS and MPEG-21 DIA BSD (see Table 2). The context descriptions are based on a subset (or profile[9][10]) of MPEG-21 DIA Usage Environment Descriptions tools (see Table 3).

Current achievements include parsing and validation of media descriptions; the adaptation decision phase and the media parameter selection are under development. At this moment, we have focused on the configuration and management of some parameters based on the constraints set for terminal presentation capabilities and network characteristics, and for the content metadata information. The user preferences are used as the starting point for parameter selection. If no user preferences are set the media information is used as default user preference. These preferences are filtered and modified in order to get a set of adaptation parameters that match the terminal and network capabilities while trying to minimize the changes on user preferences. This set of parameters is passed to the selected CAT (currently the CAT selection is based on file format) which tries to match them with it's coding capabilities in order to perform the adaptation. In those cases in which the CAT is not able to match the adaptation parameters it applies a default adaptation.

Adaptation type	Content description
Media adaptation	MPEG-7 Media Description (MediaInformation, Transcoding Hintes)
	MPEG-7 Variations and Summaries
Semantic adaptation	Regions of interest with importance (annotated by users or automatically)
Transmoding to text	MPEG-7 textual tools: Keywords, textual annotations, Spoken Content
Bitstream adaptation	MPEG-21 BSD/gBSD. Preferably gBSD to allow semantic bitstream adaptation

Table 2: Content descriptions for adaptation

Usage Environment description tools (MPEG-21 DIA)
User Description Tools
Usage Preferences
Media Format: content, bit rate, visual coding (format, frame height, frame width, frame aspect ratio and frame rate), audio coding (format, audio channels, sample rate, bits per sample).
Presentation Preferences
o AudioPresentationPreferences: Volume, output device, balance. o DisplayPresentationPreferences: Color temperature, brightness, saturation, contrast. o ConversionPreferences: Media type conversion preferences and priorities. o PresentationPriorityPreferences: Modality (audio, video...) priorities
Terminal Capabilities Tools
Codec Capabilities
Audio, video and image coding/decoding supported formats.
Display Capabilities
Supported display modes (resolution, refresh rate), screen size, color bit depth.
Audio Output Capabilities
Supported audio modes (sampling frequency, bits per sample), low frequency, high frequency, number of channels…
Storage Characteristics
Input transfer rate, output transfer rate, size, writable.
Network Characteristics Tools
Network Capability
maximum capacity and minimum guaranteed.

Table 3: Context descriptions for adaptation

Future work will focus on adapting the content using other MPEG-7/21 metadata and based on the internally extracted content features.

3.3 Content Adaptation Tools

The initial approach is to use available transcoders, in case of dealing with non scalable nor semantic adaptable content (limited by resource availability) , and to implement scalable media adaptation tools (customizing the MPEG-21 DIA Bitstream Syntax Descritpion –BSD- tools) and semantic driven media adaptation tools for non scalable content. Transmoding processes such as video to audio conversion or speech recognition would be applied whenever the user preferences indicate it or when it's impossible to adapt to the terminal capabilities in another way.
Currently, a Transcoder CAT, a Scalable Content CAT and a preliminary Real-time services CAT are integrated and operational. Future versions will also consider transmoding CATs (one "media to text" if, as expected, the corresponding metadata -mainly MPEG-7 textual description tools and Spoken Content description- are available, and another "video to slide show" [11]).

3.3.1 Transcoder CAT

The adaptation module supports initially transcoding in the pixel domain, using a full decode-encode model, setting frame rate, frame size, sampling and other coding parameters. Currently, the transcoder CAT is able to deal with MPEG-1/2/4 SP video and MPEG-1 layer 2, MPEG-1 layer 3 and MPEG-4 AAC audio. Image adaptation is also available for JPEG, GIF, … transcoding. This CAT is based on the use of ffmpeg[12].

3.3.2 Scalable Content CAT

The usage of MPEG-21 DIA's BSD tool will allow the format-agnostic adaptation of scalable content. This tool provides mechanisms to perform scalabe content adaptation via media file truncation without prior knowledge of the media file structure. This is based on the availability of xsd and xsl descriptions of the media format and media transformations respectively. Currently, JPEG2000 adaptation is implemented using MPEG-21 DIA's BSD tools, allowing resolution and quality reduction of the images. JPEG2000 adapation via transcoding is also planned to be included. Integration of the aceMedia Scalable Video will start after the SVC format is fixed.

3.3.3 Semantic driven Real-time services CAT

This module operates with compressed video. It intends to provide content adaptation according to semantic features extracted in real time from the incoming content. In order to perform the required content analysis in real-time, this module works over MPEG compressed domain parameters, such as DCT coefficients and coding motion vectors.

First prototypes will focus on temporal features (e.g., key frame extraction) and on basic spatial features (e.g., coarse spatial segmentation). These features will be then used to configure a video generation tool on the fly.

4 CONCLUSIONS

CAIN is a Content Adaptation Module target to the integration of different content adaptation approaches. The content adaptation starts with a common phase where the adaptation parameters are selected taking into account the media description and context description inputs. Afterwards, the most appropriate content adaptation tool is selected, taking into account their provided functionalities (adaptation capabilities, input and output media formats, …). Key-technologies in the operation of CAIN are MPEG-7 and MPEG-21, besides the coding formats (mainly, MPEG and JPEG families, and the scalable video format under development within aceMedia) and the integrated Content Adaptation Tools.
The proposed Content Adaptation Module will be used for service prototyping (and deployment) and as a framework for benchmarking of different content adaptation tools. Preliminary results prove that the proposed approach is feasible and that the possibility of selecting different adaptation approaches allows more flexibility depending on the usage scenario

5 REFERENCES

[1] A. Vetro, C. Christopoulos, T. Ebrahimi (eds.), "Universal Multimedia Access (special issue), IEEE Signal Processing Magazine, 20(2), March 2003.

[2] F. Pereira, I. Burnett, "Universal Multimedia Experiences for Tomorrow", IEEE Signal Processing Magazine, 20(2):63-73, March 2003.

[3] I. Komptasaris, Y. Avrithis, P Hobson and M.G. Strintzis, "Integrating Knowledge, Semantics and Content for User-centred Intelligent Media Services: the aceMedia Project", Proc. of WIAMIS'2004, Lisboa, April 2004.

[4] A. Vetro, "Transcoding, Scalable Coding and Standardized Metadata", in Visual Content

Processing and Representation-VLBV03, LNCS Vol. 2849, Springer-Verlag, 2003, pp.15-16

[5] J.R. Smith, “Semantic Universal Multimedia Access”, in Visual Content Processing and Representation-VLBV03, LNCS Vol. 2849, Springer-Verlag, 2003, pp.13-14

[6] ISO/IEC 15938-5, Information Technology – Multimedia Content Description Interface – Part 5: Multimedia Description Schemes

[7] ISO/IEC 21000-7, Information Technology – Multimedia Framework – part 7: Digital Item Adaptation

[8] M.Mrak, G. C. K. Abhayaratne, and E. Izquierdo, "On the infuence of motion vector precision limiting in scalable video coding," in Proc. Int'l conf. Signal Processing, vol. 2, Aug. 2004, pp. 1143-1146.

[9] J.M. Martínez, J.Bescós, V. Valdés, L. Herranz, “Integrating Metadata-driven content adaptation approaches”, Proc. of EWIMT’2004, London, November 2004.

[10] J.M. Martínez, V. Valdés, L. Herranz, J. Bescós, “A Simple Profile for MPEG-21 Usage Environment description tools”, Doc. ISO/MPEG M11239, MPEG Palma de Mallorca Meeting, Octubre 2004.

[11] M. Padilla, J.M. Martínez, L. Herranz, “Universal and personalized access to video summaries via multimedia presentations in the DYMAS system”, Proc. of AMR04, Valencia, August 2004

[12] http://ffmpeg.sourceforge.net/