

HYPERLINKED VIDEO WITH MOVING OBJECTS IN DIGITAL TELEVISION

*B. Cardoso¹, F. de Carvalho¹, L. Carvalho⁵, G. Fernández², P. Gouveia⁵, B. Huet³, J. Jiten³, A. López²,
B. Merialdo³, A. Navarro⁵, H. Neuschmied⁴, M. Noé², R. Salgado¹, G. Thallinger⁴*

¹ PT Inovação, S.A., Aveiro, Portugal

² La Salle School of Engineering, Ramon Llull University, Barcelona, Spain

³ Institut Eurécom, Sophia Antipolis, France

⁴ Joanneum Research GmbH, Graz, Austria

⁵ Telecommunications Institute, Aveiro University, Portugal

ABSTRACT

The GMF4iTV project (Generic Media Framework for Interactive Television) is an IST European project that developed an end-to-end broadcasting platform providing interactivity on heterogeneous multimedia devices such as Set-Top-Boxes, PCs and PDAs according to the Multimedia Home Platform (MHP) part of the DVB standard. The developed platform allows the content providers to create enhanced audiovisual contents with a degree of interactivity at moving object level or shot changes in a video. The end user is then able to interact with moving objects from the video or individual shots allowing the enjoyment of additional contents associated to them (MHP applications, HTML pages, JPEG, MPEG-4 files, ...).

1. INTRODUCTION

Hypertext is now of common use with the World Wide Web. Although the possibilities of hypervideo have already been demonstrated in several projects, its usage remains limited, in great part because of the lack of appropriate platforms for creation, distribution and presentation. For example, the Hypercafé project [1] is a nice illustration and application of a hypervideo engine. Bove et al [2] have shown how to use video segmentation and tracking to facilitate the authoring of hypervideo. The Viper system [3] allows creating personalized programs through the selection of clips at the user side.

In the Generic Media Framework for Interactive Television (GMF4iTV) project, our objective is to extend the capability of the Multimedia Home Platform (DVB-MHP) [4] standard to support personalized hypervideo in an end-to-end digital broadcasting platform. DVB-MHP specifies an open standard API (Application Program Interface), which facilitates services across broadcast, telecommunications and computer platforms and is supported on several available set-top boxes. In the GMF4iTV project, a regular TV broadcast (MPEG-2

encoded) is augmented with additional information (MPEG-7 encoded) which defines active objects in the video, along with additional content to be displayed when those objects are selected. The platform allows content and service providers to create, manage, synchronize and distribute pre-recorded linear video streams (MPEG-2/-4) in conjunction with non-linear additional content (e.g. HTML, MPEG-4, JPG) employing enhanced metadata schemes, included in an environment of live video feeds. The direct interaction with objects on the TV-screen and PDA enables the typical more lean-back oriented TV-viewer to access the additional content in an active but highly convenient way. The production part of the platform allows the development of new scenarios for a variety of program types (documentaries, sports ...).

2. GMF4iTV SYSTEM ARCHITECTURE

The complete GFM4iTV system architecture is shown in Figure 1. The edition process of the GMF4iTV platform is performed with the help of an application called GMF Authoring Tool. This tool allows the definition of objects on the screen and the automatic tracking by means of advanced video analysis algorithms. Furthermore, the user can associate additional contents such as images, HTML files, MPEG-4 videos or MHP applications to the previously defined objects or scenes. In general, the system allows the transmission of any kind of extra content if the STB is able to visualize it. Personalization of the content is also possible, the system defines several user profiles so that specific additional contents are sent to different user profiles (according for instance to their needs, expertise or age).

At the end of this process, the system produces a set of files that are sent to the multiplexer to be transmitted at the time points precisely defined in the Command Description Language (CDL) file, also generated by the Authoring Tool. At the output, a standard DVB Transport Stream (TS) is produced ready to be included in the broadcasting platform of any digital television broadcaster

either satellite, cable or terrestrial. The implemented multiplexer is also able to seamlessly switch/toggle between pre-produced interactive content and real-time encoded live video.

At the reception side, a STB or PC decodes the enriched DVB signal and presents the interactive service to the user with the possibility to retrieve the additional contents linked to objects or video shots. Complementary, the user can use a PDA, permitting to share the service within several users simultaneously, each one interacting with different objects or different additional contents.

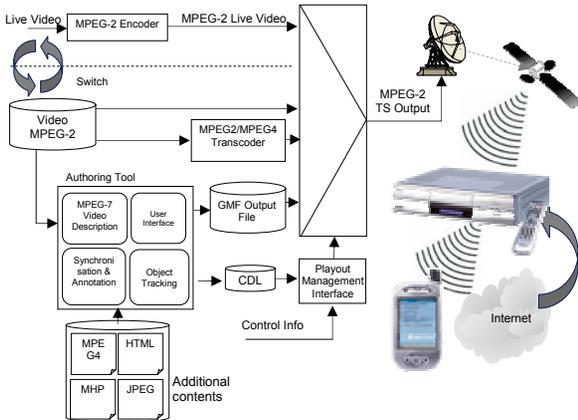


Figure 1. GMF4iTV architecture block diagram

3. GMF4iTV AUTHORIZING TOOL

The first step in the production of interactive content is the specification of objects for which additional content should be made available in the video. These objects are marked by regions. This might be a very time consuming process, therefore it is organized in the following two steps:

- Automatic extraction of the structural description of the video: shots and keyframes. Manual correction of the automatic extracted information.
- Manual specification of the interactive regions for at least one frame of a shot. Automatic tracking of the specified regions to get the region locations in all frames.

For the above listed tasks two applications have been implemented. The Import Tool allows manually entering general metadata about the video like creation and production information and information about tape location or right holders. Then the automatic analysis process for the extraction of shot boundaries and keyframes is started. The result is an MPEG-7 description (following the Detailed Audiovisual MPEG-7 profile described in [5]) for each video processed. This description is later on used as basis for the generation of the interactive region descriptions by the GMF4iTV Authoring Tool.

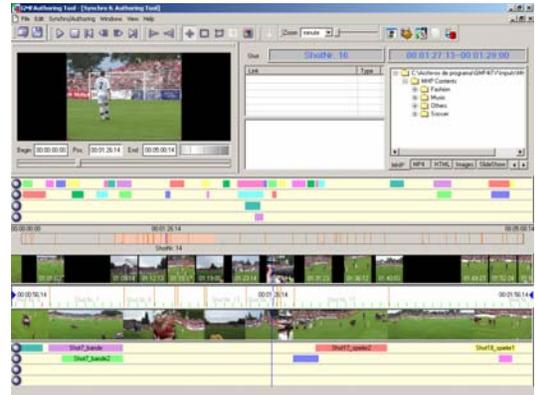


Figure 2. User interface of the Authoring Tool

The GMF4iTV Authoring Tool (shown in Figure 2) is the central tool at the production side. The video structure is shown by displaying the shot boundaries in a timeline. The integrated video player has drawing functionalities for specifying the interactive regions. The region tracking functionality of this tool produces metadata about the location of the regions in each frame. A number of views are offered to enable fast end easy navigation in the video; e.g. the object view giving information on tracked objects. These software components have been implemented by using the MPEG-7 library (freely available from Joanneum Research [6]).

The Synchronisation & Annotation Tool is a software component included in the GMF Authoring Tool which offers the functionality to associate additional contents to shots or objects in the video sequence. It also takes care that, during the edition process the available bitrate is not exceeded. Finally, it is responsible to encapsulate and synchronise all the content and corresponding metadata as described in Section 4.

Annotation of personalization information is also performed during the authoring step. For a given TV program (or program genre), three ontologies are defined: one for objects, one for shots and one for additional content. These ontologies contain semantic classes that are used to define some properties of the corresponding objects. During authoring, the user may associate one or several semantic classes to the objects, shots and additional contents of the program. Furthermore, a set of rules is defined to drive the personalization engine on the set-top box [7]. This information is encoded in MPEG-7 and multiplexed with the video data.

4. CONTENT DELIVERY

The main elements of a transport stream containing the interactive services of GMF4iTV are the following:

- **Video streams:** The main video is coded following the MPEG-2 standard. A MPEG-4 transcoded version

is also sent to the STB used to visualize the video in the PDAs

- **MPEG-7 metadata:** The system generates MPEG-7 metadata containing the video description, object definition, object position at every frame and related additional contents and personalization data.
- **Additional content:** The additional content needs to be transmitted to the STB through the broadcasting delivery chain, the transport protocol is DSM-CC Object Carousel [8].
- **NPT time base:** Finally, at reception side an auxiliary time base is required in order to resolve all the temporal references present in the MPEG-7 metadata at frame level.

4.1 MPEG4 transcoding

In order to produce the MPEG-4 video to be displayed at the PDA, a transcoding process basically consisting of transforming MPEG-2 MP@ML (Main Profile at Main Level) to MPEG-4 Simple Profile is performed at the playout, within the authoring tool. In this conversion, which should be computationally fast, we change the bit rate, frame size and frame rate. The target output parameters for this process are the following: MPEG-4 bitrate of 256 kbps; Frame rate equal to 1/3 of the MPEG-2 input frame rate (25 Hz); MPEG-4 picture size of 240x192 pixels, i.e., a reduction of 1/3 in horizontal and vertical. The transcoder output is encapsulated according to the real time protocol (RTP) and multiprotocol encapsulation (MPE). The output stream is synchronized (stamped) with the input transport stream. The transcoder architecture achieved a computational complexity reduction of 63 % relatively to a full encoding/decoding solution.

4.2 Metadata

In order to send the MPEG-7 metadata in an efficient way, the metadata is segmented before being transmitted. This metadata segmentation process is done via Access Units - Fragment Update mechanism defined in part 1 of ISO/IEC 15938 [9]. The original metadata is segmented in a common part that is sent via DSM-CC OC and a number of dynamic parts sent via Metadata Sections [10,11] (similar to private sections) that are associated to specific frames or groups of frames in the audiovisual content.

In order to unambiguously recover the time references in the metadata, the system needs a different timeline from System Time Clock (STC) provided by MPEG-2. Normal Play Time (NPT) is a continuous timeline over the duration of an event. The NPT provides an absolute timeline to which references can be made. To allow the recovery of the NPT in the receiver, the NPT reference descriptor defined in ISO/IEC 13818-6 must be sent at a

rate higher than 1Hz. This descriptor contains a relationship between the STC and the NPT.

4.3 Additional content

During the association of additional contents in the Authoring Tool, the user requires instantaneous feedback from the system about the viability of the association since the system is based on a broadcast delivery channel with a limited bitrate. This feedback is provided in an automated and transparent way from the user point of view.

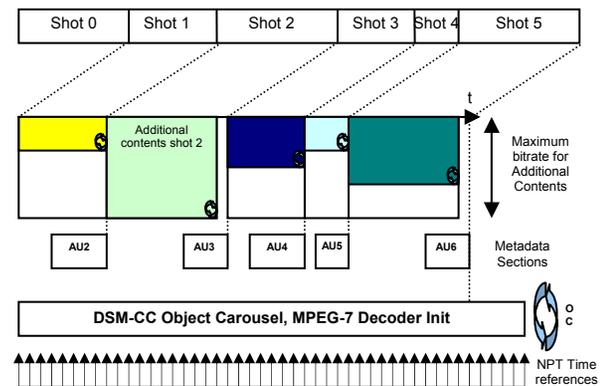


Figure 3. Content and metadata delivery

The additional content and MHP applications need to be transmitted synchronously to the object appearances and video shots to which they are related in order to use the bandwidth in an efficient way. The transmission mechanism adopted for this transmission is the DSM-CC Object Carousel, using a different carousel for each additional content (or sets of contents related to the same object or shot). Figure 3 illustrates the metadata and additional content transmission procedure. NPT time references are transmitted continuously and metadata access units corresponding to every shot are transmitted before the transmission of the associated additional content. The dynamic transmission of metadata and additional content provide an efficient use of the available broadcasting bitrate [12].

5. USER INTERFACE

5.1 Set-top-box

At the receiving side, the user interface is a MHP based application that deals with metadata synchronisation, additional content download and object highlighting. The set-top-box also multicasts video streams and metadata to PDAs.

The set-top box software includes the personalization engine which compares the semantic classes associated to the objects, shots and additional content, with the current

user profile and performs two functions: to decide if an object is interactive or not, and to select which additional content should be displayed. The user profile resides in the set-top box storage. The semantic classes and the decision rules are transmitted within the metadata associated to the program. The engine is a first-order inference engine, so that the rules can implement very complex decisions.

5.2 PDA

From the user interface point of view, the implementation of interaction based on moving objects is a challenge in itself, particularly in an Interactive TV scenario. In order to allow the full exploitation of personalized object based interaction, the GMF4iTV project introduced an innovative approach based on common Personal Digital Assistants (PDA): the PDA is used as an advanced intelligent peripheral which can be used to interact with the system, by clicking directly on the moving objects being highlighted over a video stream (synchronized copy of the main video stream shown on the TV). Furthermore, the device can be used to retrieve and display additional content related with the selected object. The video shown on the PDA is the MPEG-4 stream transcoded on the production side, encapsulated in RTP protocol and sent from the Set-Top Box (STB) over the air, using IEEE 802.11b (WiFi) technology. An additional UDP stream is generated on the STB, based on the MPEG-7 metadata and personalization options, to convey the information needed for object highlighting, synchronization and access to additional content. Upon object selection, using the conventional PDA stylus, the associated additional video, audio, graphical and text documents are retrieved from the STB, over WiFi, and displayed using the PDA web browser.

6. CONCLUSIONS

In this paper, we have described the GMF4iTV prototype which enables to provide interactive objects for TV programs. The prototype contains a complete production chain, from the identification and annotation of objects, link to additional content, semantic annotation, MPEG-7 encoding, multiplexing with the MPEG-2 video stream, decoding on the set-top box and interaction with the user through a PDA. Personalization is possible both for object activation and additional content selection.

This project is funded by the European Union under contract IST-2001-34861.

7. REFERENCES

- [1] Sawhney, N.; Balcom, D.; Smith, I.; Authoring and navigating video in space and time, *IEEE Multimedia*, Volume: 4, Issue: 4, Oct.-Dec. 1997, Pages:30 – 39
- [2] V. Michael Bove, Jr., Jonathan Dakss, Stefan Agamanolis, Edmond Chalom, Adding Hyperlinks to Digital Television, *Proc. SMPTE 140th Technical Conference*, 1998.
- [3] Stefan Agamanolis and V. Michael Bove, Jr., Viper: A Framework for Responsive Television, *IEEE MultiMedia*, vol. 10, no. 1, July - September 2003, pp. 88 – 98
- [4] ETSI TS 102 812 V1.2.1: Digital Video Broadcast (DVB), *Multimedia Home Platform (MHP) Specification 1.1.1*
- [5] W. Bailer, P. Schallauer, M. Hausenblas, G. Thallinger, MPEG-7 Based Description Infrastructure for an Audiovisual Content Analysis and Retrieval System, *Proc. Conference on Storage and Retrieval Methods and Applications for Multimedia, IS&T/SPIE Electronic Imaging*, San Jose, CA, USA, Jan. 2005
- [6] MPEG-7 library: <http://iis.joanneum.at/MPEG-7>
- [7] B. Cardoso, F. de Carvalho, G. Fernández, B. Huet, J. Jiten, A. Lopez, B. Merialdo, H. Neuschmied, M. Noe, D. Serras Pereira, G. Thallinger, Personalization of Interactive Objects in the GMF4iTV project, *TV'04: the 4th Workshop on Personalization in Future TV*, Eindhoven, 23 August 2004
- [8] ISO/IEC 13818-6: Generic coding of moving pictures and associated audio information. Part 6: Extensions for DSM-CC.
- [9] ISO/IEC 15938: Information technology – Multimedia Content Description Interface.
- [10] Generic coding of moving pictures and audio: Systems. Amendment 1: Carriage of metadata over ITU-T Rec H.222.0 | ISO/IEC 13818-1 streams.
- [11] A. López, D. González, J. Fabregat, A. Puig, J. Mas, M. Noé, E. Villalón, F. Enrich and G. Fernández, Synchronized MPEG-7 metadata broadcasting over DVB networks in an MHP application framework, *IBC'2003*, Amsterdam, September 2003.
- [12] A. López, M. Noé and G. Fernández, Linkage of additional contents to moving objects and video shots in a generic media framework for interactive television, *SPIE Conference on Multimedia Systems and Applications VII*, Philadelphia, Pennsylvania, USA, 25–28 October 2004