Personalization of Interactive Objects in the GMF4iTV project

Bernardo Cardoso[1], Fausto de Carvalho[1], Gabriel Fernàndez[2], Benoit Huet[3], Joakim Jiten[3], Alejandro Lopez[2], Bernard Merialdo[3], Helmut Neuschmied[4], Miquel Noe[2], David Serras Pereira[5], Georg Thallinger[4]

[1] PT Inovação, S.A., Aveiro, Portugal
{bernardo, cfausto}@ptinovacao.pt

[2] La Salle School of Engineering, Ramon Llull University, Barcelona, Spain
gabrielf@salleURL.edu

[3] Institut Eurécom, Sophia Antipolis, France
{huet, jiten, merialdo}eurecom.fr

[4] Joanneum Research GmbH, Graz, Austria
helmut.neuschmied@joanneum.at

[5] Duvideo, Lisboa, Portugal
dpereira@duvideo.pt

http://www.gmf4itv.org/

Abstract. This paper describes the personalization mechanism that is being implemented in the GMF4iTV Project. GMF4iTV is a European project which aims at developing techniques to create television programs with interactive objects using the DVB-MHP platform. Interactive objects are moving areas of the TV picture (generally superposed to characters, objects in the scene, or specific spots) with associated metadata. When the user selects one of these objects, the associated metadata is displayed. Personalization is used for both object and associated metadata selection. This paper describes the current version of the GMF4iTV prototype.

1 Introduction

Hypertext is now of common use with the World Wide Web. Although the interest for hypervideo has 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 [15] have shown how to use video segmentation and tracking to facilitate the authoring of hypervideo. The Viper system [14] allows creating personalized programs through the selection of clips at the user side.

In the GMF4iTV project, our objective is to extend the capability of the Multimedia Home Platform (DVB-MHP) [16] standard to support personalized hypervideo. The

DVB-MHP specification is 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 GMF4iTV project addresses all the aspects involved in the creation, distribution and presentation of personalized hypervideo programs:

- video producer side: tools to select and track video objects, and associate additional information (HTML, MPEG-4 clips, MHP applications) to those objects using an MPEG-7 structure.
- broadcaster side: multiplexing of the MPEG-2 video with the additional MPEG-7content and the MPEG-7 description, synchronization so that the required information is available when needed.
- user side: storing of the additional content, interaction with the user to select among active objects, presentation of the additional information when selected, and personalization based on user profile and preferences.

The GMF4iTV project develops a new authoring, multiplexing, server and multimedia terminal architecture for application development, synchronization, scheduling and end-user terminal systems. The GMF4iTV prototype is an end-to-end platform which provides interactivity at the moving object level. 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 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 (documentation, sports …).

The GMF4iTV prototype allows the personalization of the interaction between the user and moving objects. All objects, shots and additional content items can be labeled with specific MPEG-7 metadata, which is broadcasted together with the main program. This information is added by the video producer on the production site. On the receiver side this information is compared with the user profile by a personalization engine which uses program specific rules to activate or de-activate objects, and select the proper additional content items to be displayed.

The project makes use of advanced techniques for video coding (MPEG-2, MPEG-4), metadata encoding (MPEG-7), semantic modeling through ontologies, efficient rule processing using an inference engine.

The rest of the paper is organized as follows. We first present the various aspects of personalization within GMF4iTV, including some scenarios of interactive programs for illustration. Then we describe the authoring tool which is used to attach metadata to objects, shots and additional content items, the multiplexing tool which broadcasts the video program and the additional information together, and finally the user interaction using a PDA device.

2 Personalization in GMF4iTV

2.1 Demonstration scenarios

The following examples have been constructed in cooperation with video producers within the GMF4iTV project to demonstrate the capabilities of the prototype. Their description should give an intuitive idea of the potential of interactivity and personalization in GMF4ITV.

- **Fashion show**

  In a fashion show, models walk along a scene to present various fashion items. In the fashion show scenario, the moving objects are the models, and the additional content consists of information about the fashion items. Model objects may have the attribute Male or Female. Additional information might be about shopping possibilities or the designer.

  The user profile (which resides in the set-top box) indicates whether the user is male or female. At the beginning of the TV program, the user is asked whether he/she prefers shopping or designer information to be displayed (this choice is kept in the transient user profile). The rules of the personalization engine indicate that male (resp. female) models should be activated if the user is male (resp. female). If the user selects an activated object, the personalization engine looks at the transient profile to display the adequate shopping or designer information.

- **Music program**

The Music program shows musicians playing a sequence of songs. The additional content consists of the lyrics of the songs, information on the artists' discography and several lists of their concerts to come, organized by geographical region.

Since the discography, list of concerts and lyrics are not related to any particular video object they are simply represented as icons on the screen. The appearance of these indicators is decided during production, where the producer decides to relate certain additional content to a scene or to a time reference. The additional content is activated by user selection. In the case of the lists of concerts, the user profile contains some regional information (for example the city or zip code where the user lives) and the personalization rules contain the necessary knowledge to map this information with the geographical regions of the concerts. Therefore, only the appropriate regional list is displayed to the user.

- **Animal documentary**

The Animal documentary shows several animals in their living environment. The moving objects are the animals. Two types of additional content are provided: the adult version is a HTML text which provides further information on the animal behavior; the child version is a quiz with simple questions about the current animal. The user profile contains age information about the user, and the personalization engine uses this information to select which version of additional content to display.

The pictures below show a simulation of those two versions.

2.2 Global view on the interactivity process

The interactivity process in the GMF4iTV system can be divided into three different steps:

- on the production side, objects are defined in the video, additional content is provided, and metadata for objects and content is added,
- for distribution, the extra information for interactivity and personalization is encapsulated in MPEG-7, and multiplexed with the original TV program,
- on the user side, the set-top box retrieves the extra information, displays object locations, prepares the available additional content, and runs the personalization engine. Depending on user interaction, the adequate additional content is displayed.

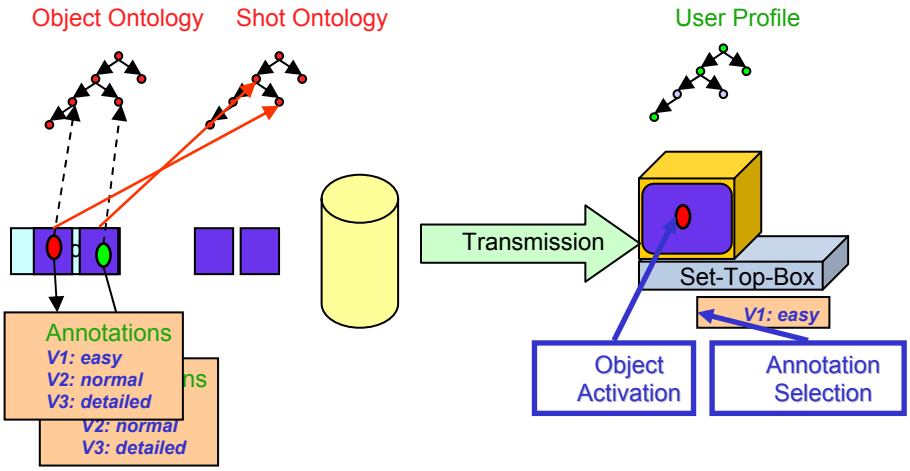


Fig. 1 This picture gives a global overview of the personalization process

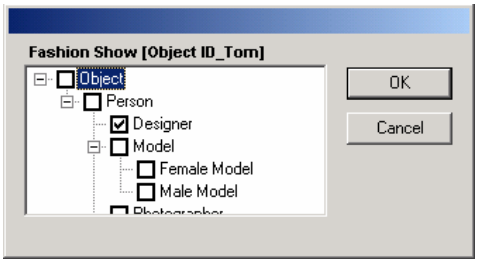
2.3 Formal concepts for personalization

Ontologies

Personalization involves taking decisions based on the nature of various items. In our system, these items are of three different types:

- Objects moving in the video (corresponding to screen areas defined by the video producer),
- Shots in which those objects appear,
- Additional content that will be displayed by user interaction.

The nature of these items is indicated by semantic attributes. The possible values of these attributes are arranged in an ontology [9][17]. Therefore, there are three ontologies for each broadcast program scenario: the object, the shot and the additional content ontology. Each ontology is a tree structure which describes the dependencies between attribute values, from the most general (root) to the most specific (leaves). The advantage of using ontologies is to give the producer a greater flexibility in the level of details during the annotation process. The figure below shows an example of selection of an attribute within an ontology.



Note that in the applications that we envision for the GMF4iTV prototype, scene and additional content ontologies are generally simplistic, the object ontology being more elaborate.

After discussion, we have decided that those ontologies would be scenario specific (although an ontology can be reused in other scenarios). This avoids the problem of constructing a "universal" ontology which would be known by both the producer and the user sides. As a result, scenario ontologies are encoded in MPEG-7 and transmitted to the user side at the beginning of the program.

User profile

Information about the user is kept in the user profile, which resides in the set-top box on the user side. The user profile is split in two parts a static one (static profile) and a transient one (transient profile). The static profile is not scenario dependent. For the time being, it is manually built by the system administrator, although in real situations, it should be created by the user herself.

In several scenarios, there is the need for some extra information from the user. This may be because this information is not in the user profile (it cannot contain every possible information about what he likes or dislikes), but also because some information is not static (for example it might depend on the mood of the user). Therefore, a

GMF4iTV program may start with a small sequence of questions which allows the user to make custom choices. The answers to those questions are kept in the transient profile which remains valid as long as the scenario is broadcasted, and erased afterwards.

Personalization engine

The personalization engine has two functions:

- Based on the comparison between the attribute of objects and scenes and the user profiles, decides which objects can be activated for this user,
- Based on the above plus the additional content attributes, decides which version of additional content should be displayed.

For maximum flexibility, the personalization rules are written as inference rules, and the personalization engine is a first order inference engine. Attributes and user profile characteristics are facts. Running the inference engine on those facts will start an inference chain which will eventually trigger an activation or selection predicate. The personalization rules are also encoded in MPEG-7 and transmitted at the beginning of the program.

Usages of personalization

The mechanisms described above are quite general and can be used in a great variety of situations. Extensive discussions with video producers have generated a number of potential usages for these mechanisms, of which the following is just an illustrative sample:

- Selection based on user type: male users would like to have information on male clothes, female users on female clothes,
- Selection based on age group: adult and child version of additional content,
- Comparison of characteristics: geographical information,
- Language dependent information
- Etc…

3 MPEG-7 Annotation

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. This objects marked by regions. This might be a very time consuming process, therefore it is organized in the following steps:

- Automatic extraction of the structural description of the video: shots and key-frames.
- 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.

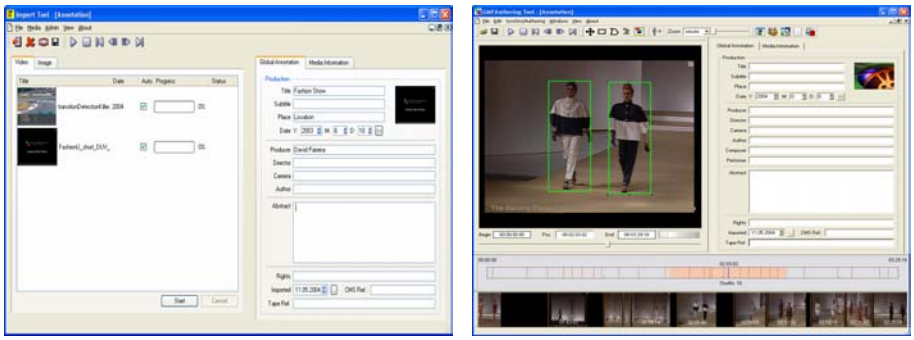


Fig. 2 Import Tool (left image) and GMF4iTV Authoring Tool (right image) used for video annotation

For the above listed tasks two applications have been implemented (see Fig. 2). The Import Tool allows to manually enter 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 key-frames is started. The Import Tool supports batch mode so it is possible to handle more than one video at a time. The result is an MPEG-7 description for each video processed. This may be stored in a database and later on used as basis for the generation of the interactive region descriptions by the GMF4iTV Authoring Tool.
The GMF4iTV Authoring Tool provides a user interface for easy navigation in the video. The video structure is shown by displaying the shot boundaries in a timeline. These shot boundaries can be edited for manual correction. 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. These software components have been implemented by using the MPEG-7 library freely available from Joanneum Research [7].

4 Authoring and Multiplexing

In order to associate additional contents to objects or shots within a digital video sequence, Authoring and Multiplexing tools are required. The Synchronization and Authoring tool provides the user interface to associate additional content on one side, and the data encapsulation and synchronization on the other. The process produces a script file (conformant to a Content Description Language, CDL) that is delivered to the Multiplexing tool. This module is responsible for accurately multiplexing all the previously produced streams.

4.1 Synchronization & Authoring Tool

The Synchronization & Authoring Tool is a software component included in the GMF4ITV Authoring Tool which is responsible for the following tasks:

- Management of the associations of additional contents to objects, shots or the whole video in a user friendly manner. The process is done taking into account the personalization ontologies, where every additional content item is associated with according entries from the ontologies.
- Encapsulation of a dynamic MPEG-7 metadata service over MPEG-2 according to the corresponding recent amendment [2], by the use of the Fragment Update mechanism (FU) over Object Carousel and Metadata Sections described in [3,4].
- Encapsulation of the associated additional contents such as MHP, MPEG-4, JPEG or HTML over MPEG-2 by using the DSM-CC Object Carousel protocol according to [5].
- Synchronization of MPEG-7 metadata and additional contents to the main MPEG-2 video at frame level by the use of the Normal Play Time concept, NPT [5].
- Global bitrate management.

The Synchronization & Authoring Tool interprets the results of the edition process made by the operator and optimally allocates the transmission period and bitrate for every associated content item. The process also checks the viability of every association according to the already allocated bitrate and time constraints. The complete edition process is done in a friendly and transparent manner adapted to users not familiar with DVB/MHP technology.
Finally, the Synchronization & Authoring Tool generates the CDL file containing the commands for the multiplexer. The CDL file plus the other files containing the additional encapsulated contents, MPEG-7 metadata, NPT and signaling are taken by the multiplexer to generate the output in a DVB Transport Stream (DVB-TS) compliant format.

4.2 Multiplexing

The multiplexer has as main task to multiplex the different contents referenced in the CDL at the marked time points and produce a final Transport Stream to be sent to a DVB network (satellite, cable or terrestrial). The main components of the output DVB-TS are:

- MPEG-2 Video.
- The transcoded MPEG-4 video for use in the PDA.
- DSM-CC Object Carousel and Metadata Sections containing the MPEG-7 metadata.
- Normal Play Time, for synchronization of metadata and different media.
- DSM-CC Object Carousels containing additional contents (including MHP apps).
- Application Information Table (AIT) needed for MHP signaling.

In order to combine the object-based interactive content in a real TV broadcast scenario, the multiplexer is able to seamlessly splice between MPEG-2 encoded live streams and pre-produced GMF4ITV contents.

5 User Interaction

The intrinsic temporal and spatial nature of video both complicates and invites to new navigation concepts. Assuming that the traditional TV viewer is primarily looking for passive entertainment experiences, user intervention is set to take place on a supplementary mini-screen, in order to not disturb the main picture. Keeping the interaction on a separate unit also allows us to experiment with new ways of navigation.

5.1 Interaction based on 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, 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, a 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 a MPEG-4 stream transcoded from the main MPEG-2 video stream on the production side, encapsulated in RTP protocol and sent from the Set-Top Box (STB) to the portable device 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.
A specific client application has been developed for the Windows Mobile Pocket PC used in the project. This client includes modules to extract, synchronize and render the MPEG-4 packets, as well as to overlay graphical representation of the moving objects on top of the video stream. 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.
Besides being a very intuitive and easy way of interacting with a rather complex system, the introduction of the PDA concept includes other attractive features. In particular and considering the evolution to a multicast/multi-user scenario, which is closer to the way people watch TV, it will offer the possibility to have more than one person on the same room interacting with the system without interference, because the TV main screen can be kept clean without additional elements beyond the main TV show. Another possibility will be the download of additional content to the PDA for later viewing (either push or pull).

Since all the interaction can be performed using a personal device, deep personalization can be achieved and offered as a powerful business opportunity to service and content providers, as well as real added value to final costumers. With the increasing momentum for small powerful personal devices, boosted by the introduction of UMTS and the popularity of WiFi, personalization on Interactive TV will certainly reach new exciting levels in the near future.

6 Conclusion

In this paper, we have described the GMF4iTV prototype which provides 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. The personalization process is based on rules which are processed by an inference engine on the set-top box. This powerful process allows a very flexible usage of personalization in a variety of situations provided by different application scenarios.

The project is now in the integration phase. All modules have been designed and implemented by the different partners, according to common specifications. They are currently being integrated together to compose a complete chain. A set of scenarios has been defined and according video sequences have been prepared which are being annotated with the authoring tool. A public demonstration of the project is planned at the IBC 2004 Conference.

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. *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.
3. ISO/IEC 15938: Information technology – *Multimedia Content Description Interface*.
4. A. López, D. Gonzàlez, J. Fabregat, A. Puig, J. Mas, M. Noé, E Villalón, F. Enrich and G. Fernàndez : *Synchronized MPEG-7 metadata broadcasting over DVB networks in an MHP application framework*, IBC'2003, Amsterdam, September 2003.
5. ISO/IEC 13818-6: *Generic coding of moving pictures and associated audio information*. Part 6: Extensions for DSM-CC.
6. Nack, F., Lindsay A.: *Everything You Want to Know About MPEG-7: Part I & Part II*, IEEE Multimedia, Vol. 6 No. 3 & 4, (July - September 1999) 65-77, (October - December 1999) 64-73
7. MPEG-7 library: http://iis.joanneum.at/MPEG-7

8. Naphade, M.R.; Huang, T.S., *Extracting semantics from audio-visual content: the final frontier in multimedia retrieval*, IEEE Transactions on Neural Networks, Volume: 13 Issue: 4 , July 2002 Page(s): 793 -810
9. Town, C.; Sinclair, D. *Ontological query language for content based image retrieval*, CBAIVL 2001 - IEEE Workshop on Content-Based Access of Image and Video Libraries, 2001, Page(s): 75 -80
10. Kuo, Chung-Chieh Jay, Guo, Ju, *Semantic video object segmentation for content-based multimedia applications*, Kluwer Academic Publishers, 2001, ISBN : 0792375130
11. Dai, W., Cohen, R. (2003). *Dynamic Personalized TV Recommendation System*. In-Proceedings of the 3rd Workshop on Personalization in TV. Pittsburgh, U.S.A.
12. H. Lee, J. Nam, B. Bae, M. Kim, K. Kang and J. Kim: *Personalized Contents Guide and Browsing based on User Preference*. Second International Conference on Adaptive Hypermedia and Adaptive Web Based Systems: Workshop on Personalization in Future TV, Malaga, Spain, 2001, pp.131-150.
13. L. Ardissono, F. Portis, P. Torasso. F. Bellifemine, A. Chiarotto and A. Difino . *Architecture of a system for the generation of personalized Electronic Program Guides*. Workshop on Personalization in Future TV, 2001.
14. 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
15. V. Michael Bove, Jr., Jonathan Dakss, Stefan Agamanolis, Edmond Chalom, *Adding Hyperlinks to Digital Television*, Proc. SMPTE 140th Technical Conference, 1998.
16. ETSI TS 102 812 V1.2.1: *Digital Video Broadcast (DVB), Multimedia Home Platform (MHP)* Specification 1.1.1
17. Antoniou, Grigoris, van Harmelen, Frank : *A semantic Web primer*, MIT Press, 2004, - ISBN : 0262012103