Video Summaries Generation and Access via Personalized Delivery of Multimedia Presentations Adapted to Service and Terminal

Marta Padilla,[†,§] José M. Martínez,[*] Luis Herranz[‡]
Grupo de Tratamiento de Imágenes, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Ciudad Universitaria de Cantoblanco, E-28049 Madrid, Spain

This article is centered on describing the provision of universal multimedia access services for video summaries via multimedia presentations that allow the integration of multimedia messaging service (MMS)-enabled terminals in the framework of the deferred time environment (DTE) of the DYMAS system. The system uses the framework of MPEG-7 and MPEG-21 to provide description metadata of the multimedia content and the usage context (including terminal, network capabilities, and user preferences), respectively. These descriptions are the base for the main functionalities of the complete system that provides personalized access to content (filtering by user preferences or via querying) that is first adapted to a multimedia presentation (generating a video summary that is represented via keyframes with synchronized audio clips) by the Presentation Builder and afterward adapted to the current service and terminal by the Adaptation Engine.

1. INTRODUCTION

Universal Multimedia Access (UMA)[1] refers to the capability of access to rich multimedia content through any client terminal and network. The development of new access networks, providing multimedia capabilities, and a wide and growing range of terminals makes the adaptation of content an important

*Author to whom all correspondence should be addressed: e-mail: JoseM.Martinez@uam.es.
[†]e-mail: Marta.Padilla@uam.es.
[‡]e-mail: Luis.Herranz@uam.es.
[§]Marta Padilla is currently working for Accenture (Madrid, Spain); e-mail: marta.padilla.miguel@accenture.com.

INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS, VOL. 21, 785–800 (2006)
 Published online in Wiley InterScience (www.interscience.wiley.com). • DOI 10.1002/int.20160

issue in future multimedia services, where metadata will play an important role.[2] MPEG-7[3] and MPEG-21[4] are the new MPEG standards aiming at providing the description tools supporting the creation of such metadata, in the form of content descriptions and context descriptions.

The motivation for creating the DYMAS system was the creation of new services and applications from digital television sources based on the digital video broadcasting (DVB) standards[5] that are the base for digital television in Europe (among other regions). Two kinds of services were envisioned:

- Provision of an enhanced DVB compliant stream with digital television interactive applications (following the DVB–Multimedia Home Platform [MHP] specification), either synchronized with the broadcasted content in real time (e.g., summaries of current emission) or more traditional enhanced TV applications (e.g., hooks to related information that appear depending on the contents in the broadcasted streams).
- Near real-time services over alternative services and networks (e.g., web portals, personalized services to mobile devices) where content is delivered in a personalized way (depending on the user preferences for content) and adapted to the preferred service and terminal.

In this article we review shortly the current status of this subsystem enabling UMA using both MPEG-7 and MPEG-21, detailing the multimedia messaging subsystem, which is in charge of generating video summaries and providing access to them via multimedia presentations, adapted both to the service (web portals, e-mail, MMS mobile phones) and the terminals characteristics (e.g., resolution).

Different authors have published about general issues and architectures for UMA systems,[1,6–9] but there are not so many papers about prototypes, test beds, or implementations (e.g., Refs. 10–12) and only recently have systems working with MPEG-7 and MPEG-21 descriptions been published (e.g., Refs. 13 and 14). The DYMAS system provides, besides a real-time environment for generating metadata to enable the provision of added-value interactive digital television applications synchronized with content,[15] an environment enabling the provision of alternative multimedia services based on the digital television broadcasted content.[16] The work presented in this article, besides using both MPEG-7 and MPEG-21 for, respectively, content description and context usage description as in Refs. 13 and 14, works over digital television content and aims at providing different real-time services over different terminals and networks: for Internet-based PCs to mobile phones, passing through digital television infrastructures. Another additional issue is that the generation of the video summaries is integrated in the complete personalization and adaptation framework.

The article is structured as follows. After this section introducing the motivation, objectives, and related works, Section 2 introduces the DYMAS system and overviews the current architecture of the subsystem enabling UMA. Section 3 describes the metadata framework used to enable adaptability. Section 4 describes the multimedia presentation subsystem, and Sections 5 and 6 detail, respectively, the video summary generation process and the personalized access to content functionality. Section 7 concludes the article.

2. OVERVIEW OF THE DYMAS SYSTEM

The DYMAS System (see Figure 1) is mainly a multimedia processing system with one information input, a DVB transport stream (DVB-TS), and two information outputs (the modified DVB-TS and audiovisual services directed to other alternative access networks).

The system relies on technology for automatic content extraction from audiovisual information, which is the basis for on-line service provision, that is, for the real-time environment (RTE).[15] Besides this, in order to be open and scalable and to progressively adapt to research results, the framework here presented also considers the provision of services that do not have a real-time requirement, but can conversely be offered with some delay. This is a responsibility of the deferred-time environment (DTE).

The DTE provides several UMA-related services using different types of content adaptation technologies (e.g., transcoding, semantic-based summary generation, adapted streaming services). Depending on whether it is the user who request content or it is the system who sends the content without explicit request of the

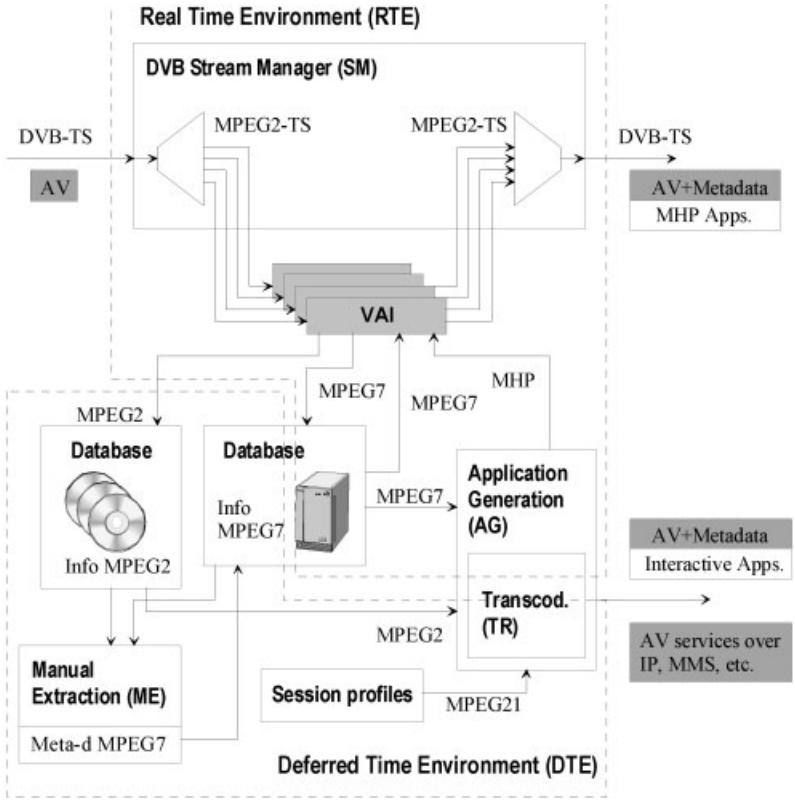

Figure 1. DYMAS system architecture.

user (according to subscription parameters), services can be pull ones (e.g., Web portals) or push ones (e.g, messaging services), where personalization is provided via transparent content filtering (enhanced with additional searching restrictions if the user wants it) in the first case and via content selection (using the preferences for the definition of the subscription) in the second one.

2.1. DTE Subsystem Architecture

The current DTE provides universal and personalized access to the MPEG-2 database via PCs on the Internet,[16] via mobile Java terminals,[17] and via MMS terminals.[18] One of the DYMAS DTE subsystems is the multimedia messaging subsystem, which allows access to video summaries (of the contents in the MPEG-2 database) in a web portal and the delivering of the summaries as multimedia presentations (using the synchronized media integration language [SMIL] standard,[19] a markup language designed to present multiple media files synchronized) to MMS terminals and e-mail clients. The generic DTE system architecture is shown in Figure 2. At the current status of the system, the applications and mobile terminal simulators run over standard PCs. The MMSC emulator and the Mail server are necessities for delivering multimedia messages to the terminal emulators and to the e-mail clients, respectively.

3. METADATA FOR ADAPTATION

A key component of the whole framework of DYMAS is the metadata, both content descriptions and the usage context descriptions. The two more recent

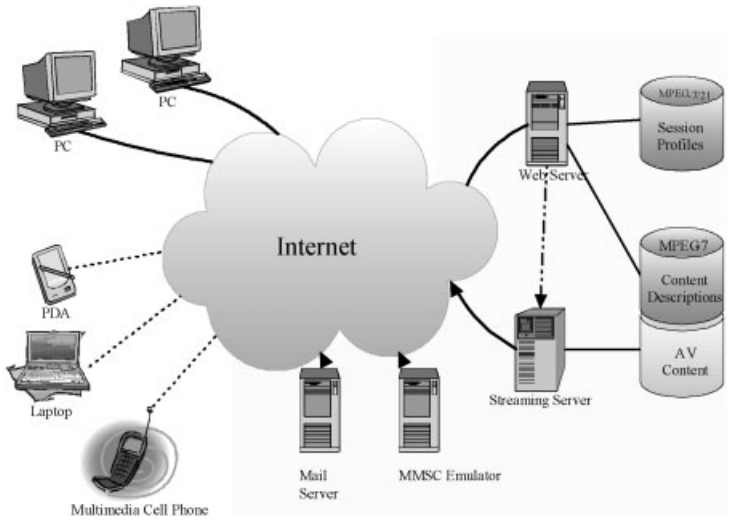

Figure 2. Overview of DTE subsystem components.

standarization efforts of MPEG tackle the need of standarized metadata to describe the multimedia content, in both high- and low-level metadata (MPEG-7), and the need of a reference framework, using other metadata that enables the provision of enhanced services, including adaptation through different usage contexts (MPEG-21).

DYMAS uses mainly the MPEG-7 multimedia description schemes (MDS)[20] to add semantic information to the MPEG-2 media. Some parts of the descriptions are generated in the RTE, and, besides their use in interactive television applications, are stored in the MPEG-7 database, where these descriptions are enriched via additional (non-real time) automatic and supervized feature extraction algorithms. Other metadata, such as genre, titles, agents (e.g., actors, anchormen, etc.), abstracts, or other high semantic level metadata related to the content can be added off-line by manual annotation. MDS is also used to describe the user preferences, which also enable a personalized retrieval of the content, selecting the content to be delivered. These preferences are used to filter the content to the user.

The framework of MPEG-21 is built on two main concepts: the digital item and users. A digital item is a structured digital object, with related metadata, and is the fundamental unit of delivery and transaction within the framework. The users are entities (such as individuals, consumers, etc.) that interact with other users, using digital items. DYMAS uses mainly the MPEG-21 digital item adaptation[21] specification, which provides description metadata of the usage context, in order to perform content adaptation to the user usage context. These metadata include:

- user characteristics, which include search and filtering preferences (imported from MPEG-7 schema) and allows filtering according to the content preferred by the user, such as genre, author, and so forth;
- network capabilities, including important values for adaptation such as the maximum bitrate allowed; and
- terminal capabilities, including display resolution and decoding capabilities.

Currently, all these metadata are stored using a common relational database management system (DBMS). Though MPEG-7 and MPEG-21 descriptions are XML documents, it is difficult to manage these kinds of data where the structure is not fixed and which is also highly typed. To simplify the management, a selected subset of MDS, covering the useful metadata for the system, has been mapped to a fixed relational structure in the DBMS. General issues about MPEG-7 data and their storage in DBMS, using database extensions or native XML databases, can be found in Ref. 22.

4. MULTIMEDIA PRESENTATION SUBSYSTEM

This part of the system (see Figure 3) consists of the following applications or modules: registration of multimedia contents, terminal, and users; advanced search of contents; SMIL/MMS presentations builder; personalized delivering of video summaries (e-mail delivering and MMS delivering simulation); presentation; and browsing.

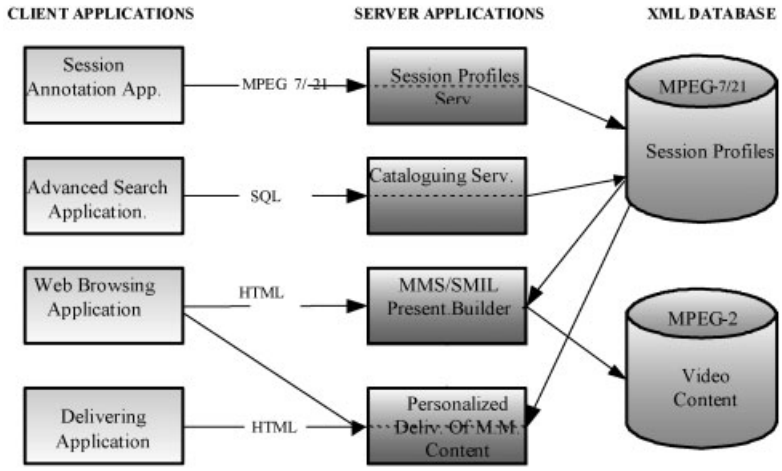

Figure 3. Multimedia presentation subsystem.

- The **annotation application** provides an easy interface to edit XML descriptions using a set of MPEG-7 and MPEG-21 description tools. These descriptions can refer to the multimedia content (according to the MPEG-7 schema) and be used to catalog new multimedia content, user preferences, or terminal capabilities (according to the MPEG-21 digital item adaptation schema, wrapped within the MPEG-21 digital item declaration schema). At the server side, new descriptions are stored in the corresponding description databases: the content description database and the session profiles database (see Figure 2). In addition, it is possible to register new contents, users, and terminals starting from MPEG-7/21 descriptions built by other applications that generate this kind of description.
- The **advanced search application** provides a query-driven search over the descriptions databases. It allows searching contents, terminals, and users. Users can build queries by specifying the value of the descriptors they are interested in by using them as search criteria. The different descriptors are joined by the logical AND operator. The server sends the search results as a HTML page to the client terminal (no relevance feedback mechanism has been currently considered within this work, as its main objective is the generation of video summaries).
- The **MMS/SMIL presentation builder application** takes charge of extracting the temporal structure, discards the irrelevant segments, and generates the multimedia material (audioclips synchronized with shots after discarding irrelevant segments and the corresponding keyframes) needed to compose afterward the SMIL presentations.
- The **web browsing application** allows users to select content among the search results page and accessing more elaborated versions of the summaries, composed with the extracted material resulting from the MMS/SMIL builder application. To play these presentations within the web application, we use a SMIL player, RealOne, which offers the capability of being embedded in HTML code. The user will only be able to access presentations of the contents of the MPEG-2 database that agree with his preferences (when creating the browsing page the contents are filtered by content description, and only those matching the user preferences are presented to the user).
- The **personalized delivery of multimedia content application** takes charge of delivering the SMIL video summaries to the e-mail accounts of users and simulating the delivering of previews to MMS enabled terminals (in fact to MMS terminal simulators). To perform this personalized delivery, the application bears in mind the content preferences

of the user described in the MPEG-21 preference description file. Users specify their preferences when they first register in the web application and can change them whenever they want. All or parts of them are used for the three different services for accessing the video summaries: web portal, mailing of summaries, and MMS to mobile terminals.

Regarding this subsystem, in the next sections we will detail the video summary generation process and the personalized access to content (either via the web portal or the messaging services).

5. VIDEO SUMMARY GENERATION

In the multimedia presentation subsystem, we provide access to video via summaries of it consisting of slide shows of images and synchronized audio clips. To generate the presentations to be accessed (either via a web portal or by delivered messages), first a master summary is built, and afterward it is adapted to the different services the application offers (web portal browsing, e-mail delivery, and MMS delivery). In the case of MMS, an additional step is performed, that is, the adaptation of the master MMS to the terminal capabilities, currently mainly display resolution. (See Figure 4.)

5.1. Video and Audio Processing

To compose the multimedia presentations, for each MPEG-2 video in the database, we perform temporal video segmentation and keyframe extraction, followed by audio processing. This process is performed automatically each time the application detects that a new MPEG-2 content has been uploaded to the content database. With all the outputs (keyframes and associated audio clips) of the signal processing modules, the master summary is composed as a multimedia presentation. (See Figure 5.)

5.1.1. Video Temporal Segmentation and Keyframe Extraction

To create the summary, we need first to extract its visual components. As only one keyframe will be selected to be synchronized with the associated audio for each "slide" in the presentation, first we need to obtain temporal intervals and then choose their associated keyframes.

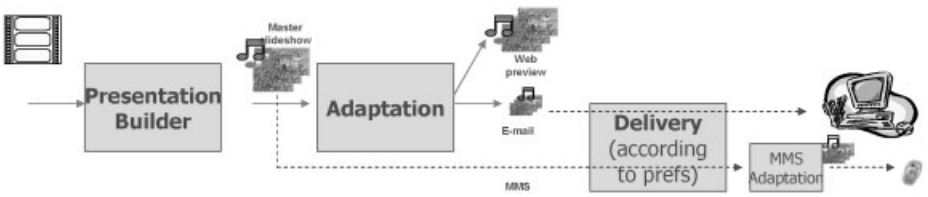

Figure 4. Video summarization architecture.

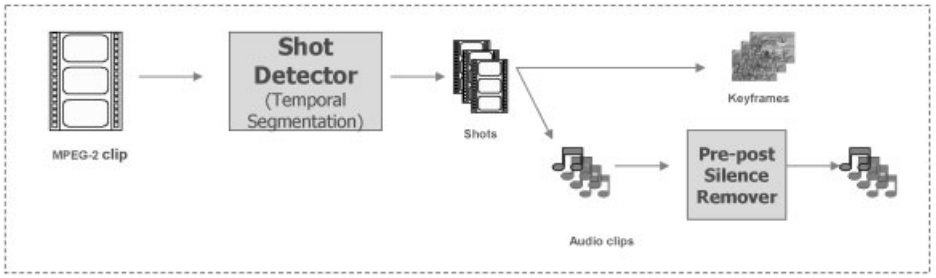

Figure 5. Video and audio processing.

To obtain these outputs (interval and keyframe), we need some information about where (in which frames) transitions of scenes take place, that is, the temporal segmentation of the video. To obtain the shot boundaries, we use an algorithm based on the application of a shot change detector model to a combined set of metrics that work with DC images that achieves results similar to detectors operating at a pixel level, namely, a recall of 99% and a precision of 95% for the most common shot changes (cuts) and 87% and 66% for gradual shot changes.[23] The average number of shots detected widely varies depending on the particular content type (musical videos, news, trailers, films, etc.), its duration, and so forth.

Many of the algorithms for shot change detection and keyframe extraction deal with divergence of frames, at the pixel level, measured with different metrics. When dealing with high resolution videos, such as digital television content (720×576 pixels, 25 frames per second), the amount of data to be processed at the pixel level (when it is decoded) is huge. Processing such data is not affordable by a real-time system. To reduce the amount of data, DC images can be used, which are spatially reduced versions of the uncompressed frame. In addition, some processing is saved avoiding the last stages of the decoder (e.g., inverse DCT, motion compensation at pixel level). They are built taking advantage of the block DCT-based coding scheme of MPEG and using the property of the DCT that the first coefficient (DC) of the transformed block is the mean of the original block.

This algorithm first partially decodes the MPEG-2 video stream to extract the DC images of each frame. As the DCT size is 8×8, that is also the reduction of the frame size, giving a DC image size of 90×72 for DVB. Then, the shot detector computes the divergence among several DC frames and decides where the shot changes are. Figures 6 and 7 show an example of an uncompressed frame and the DC image.

For each shot a keyframe is selected. The algorithm for keyframe extraction within a shot allows us to get several keyframes per shot, if necessary. This algorithm computes the divergence from the last keyframe and selects the frame as a keyframe when a threshold is exceeded.

Afterward each keyframe selected is fully decoded, having a full-size image for further adaptation.

The temporal segmentation into shots and the extracted keyframes are raw data to be further adapted to terminal and networks.

Figure 6. Frame (720 × 576) from MPEG-2 stream.

Figure 7. DC image (90 × 72).

5.1.2. Audio Processing

For each selected interval and keyframe, we extract the audio fragment associated with it (i.e., between the interval shot boundaries) and analyze it using an algorithm based on the use of a similarity criteria over an eight-features vector: overall energy, ZCR, cepstrum, roll-off, and energy of four subbands.[24] The result is a classification of the audio fragment in silence, speech and music, and noise segments.

The silence segments that commonly appear at the start and end of the audio fragments are removed, therefore saving the transmission of this null information (being multimedia presentations, the silences provide little value).

We are also working in providing feedback to the video temporal segmentation about the type of audio segment within the shot in order to increase the number of intervals (at subshot granularity) and therefore also the number of candidate keyframes that would be afterward selected depending of the maximum number of desired frames, for example, giving more importance to keyframes associated with speech or filtering based on the mood of the music (music mood is usually used for marking, either supporting or enhancing, the scene visual content).

5.2. Composition of Service-Adapted Multimedia Presentations

With the multimedia material (keyframes and associated audio clips) extracted from the analyzed video segments, we can already compose the multimedia presentations that will be accessible via the web and from the delivered e-mail messages. The process of composing the MMS SMIL presentations is explained below, as it requires one more step.

SMIL has been chosen as the representation format as it can be reproduced by most Internet browsers and e-mail clients, and because it is the base of the MMS representation format.

The system builds three versions of the multimedia presentation for each audiovisual content. Therefore the system uses three different templates, one for each service, as each of the three services has its own particular nature and offers different possibilities and limitations (display capabilities, network resources, etc.) These templates are called SMIL masters.

The SMIL master for the web presentation is the more complex and powerful one. It uses mouse events and transitions effects. From the web portal, users can access the multimedia summaries of the MPEG-2 contents consisting of a two-little-screens layout SMIL presentation; one of these screens, if clicked on, shows a slide-show presentation consisting of images and audio clips combined with transitions effects, and the other one shows the complete and original MPEG-2 multimedia content when clicked on. This presentation also has player control buttons enabled, so the user can manage the reproduction of the SMIL presentation. (See Figure 8.)

The e-mail SMIL master consists of a simplified version of the web presentation, with single-screen layout, which displays the slid show automatically when opening the e-mail. It also has player control buttons enabled. (See Figure 9.)

Finally, the MMS SMIL master is the simplest one. It is a simple slide show, without transition effects, based on SMIL 2.0 Basic Profile,[25] a subset of the W3C standard, which allows the use of SMIL 2.0 in devices with few capabilities. The SMIL presentations for the MMS scenario are built with JPEG images and AMR audio clips to minimize the size of the messages, but the standard allows the use of other codification formats for both images and audio. (See Figures 10 and 11.)

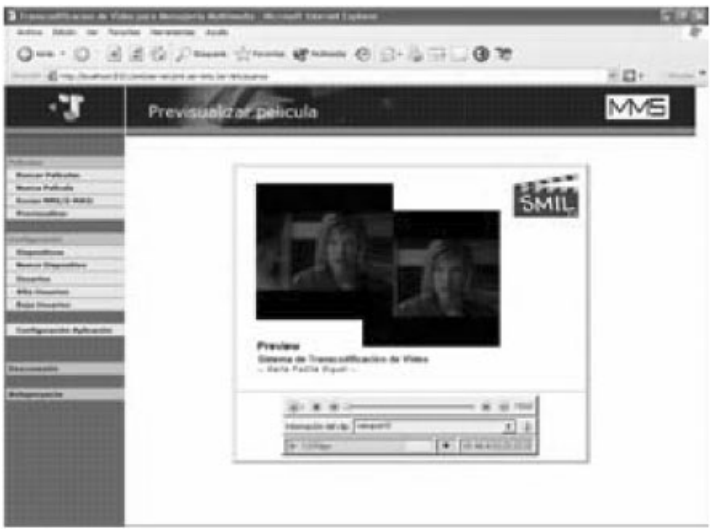

Figure 8. Web portal multimedia presentation.

Figure 9. E-mail multimedia presentation.

Figure 10. MMS multimedia presentation.

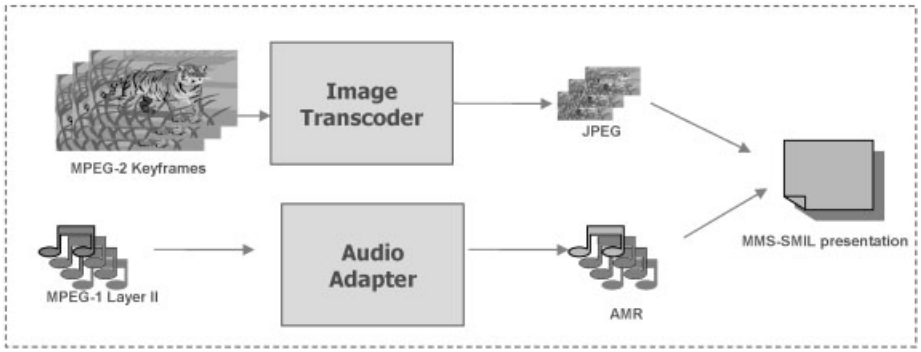

Figure 11. MMS adaptation to terminal.

5.3. MMS Adaptation to Terminal

Besides the adaptation that is performed by the transmoding of the video into a video summary in the form of a multimedia presentation (slide show), the Master MMS should be adapted each time a new terminal (with different display characteristics) requests it. This means that if the system needs to send an MMS summary of an MPEG-2 video to a model of terminal for which the presentation has already been sent, it will not be necessary to build the same presentation again, because the presentations are stored once they have been built, together with its adapted multimedia material, in order to be sent to a terminal.

The most important parameters needed to adapt the MMS are display resolution, in order to transcode the images, and eventually the decoding capabilities of the terminal. This information is stored in the terminal description (using MPEG-21 DIA).

Control parameters regarding the size of presentations have not been currently implemented. In the tests made dealing with MMS (restricting the number of keyframes up to 10, which implies an additional filtering of keyframes if the summary includes more keyframes), the size of the messages was within 15–100 KB, which is the recommended range of size to be sent via MMS systems in their first phase. The size of images is within 2–3 K and within 1–15 K for audio clips.

6. PERSONALIZED RETRIEVAL AND DELIVERY OF CONTENT

Besides the content adaptation performed for the three different scenarios (web portal, MMS, and e-mail messages) and for each terminal in the case of the MMS scenario, the system provides personalized access to content using three different set of metadata (descriptions): the content description, terminal capabilities (see Figure 12), and user preferences (see Figure 13).

The content description is based on the MPEG-7 simple profile[26] content descriptions and are created by the annotation application (see Section 4). The user preferences and terminal capabilities description uses MPEG-21 DIA description tools. Depending on the scenario, the use of these metadata can vary slightly.

6.1. Pull Services: Web Portal

Depending on how the retrieval is carried out, we can distinguish two possible scenarios. If it is the user who initiates the retrieval, querying the system for a specific content, we are in a *pull* service (the user pulls the content from the system). However, if the system sends the content to the user without an explicit query from the user, we are in a *push* service (the system pushes the content to the user).

The access to the multimedia content database from the web portal is a pull service, where the user retrieves a specific content using the advanced search application, logged with common user permissions, that is, allowed to search content. The contents available to users are filtered according to their preferences and the

```
<?xml version="1.0" encoding="UTF-8"?>
<mpeg21:DIA
xmlns:mpeg21="urn:mpeg:mpeg21:dia:schema:2003"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance"
xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001">
 <mpeg21:Description
xsi:type="mpeg21:TerminalCapabilitiesType">
  <mpeg21:InputOutput>
   <mpeg21:Display bitsPerPixel="64"
colorCapable="true">
  <mpeg21:Resolution horizontal="98" vertical="65" />
  ...
   </mpeg21:Display>
  </mpeg21:InputOutput>
  <mpeg21:DeviceProperty>
<mpeg21:DeviceClass>MobilePhone</mpeg21:DeviceClas
s>
  </mpeg21:DeviceProperty>
 </mpeg21:Description>
</mpeg21:DIA>
```

Figure 12. Example of terminal capabilities description.

annotated content descriptions; that is, they are not allowed to access contents not matching their preferences.

The delivery is carried out when the user selects one of the results of the search, using the usual HTTP protocol through the Internet.

6.2. Push Services: E-mail and MMS

DYMAS provides two push services: e-mail and MMS. In both cases, content is submitted by an authorized user (with administrator profile) to the corresponding users. Only previews of contents matching user preferences are delivered to the subscriber. The delivery application takes charge of delivering the multimedia presentation as e-mail or MMS messages to those users whose preferences match the description associated with the summarized video. Currently, no automatic policy is used to select which content matching the preferences has to be submitted. The authorized user is the one who decides when and to whom to submit the presentation.

Users have a subscription where a policy of delivery is stated (e.g., a number of e-mails/MMS per month), according to the user preferences. For instance, a user can prefer only new music content or only recent news. These genre preferences are stored in the user preferences description (using MPEG-21 DIA). The genre is annotated in each of the contents populating the database (using MPEG-7 MDS).

To perform the e-mail message delivery, we use a mail server that sends messages according to SMTP protocol. The MMS delivery is actually a simulation,

```
<?xml version="1.0" encoding="UTF-8"?>
<mpeg21:DIA ..>
 <mpeg21:Description xsi:type="mpeg21:UserCharacteristicsType">
  <mpeg21:UserInfo><mpeg7:Name>
    <mpeg7:GivenName abbrev="javi">Javier</mpeg7:GivenName>
    <mpeg7:FamilyName>Garrido</mpeg7:FamilyName>
    <mpeg7:FamilyName>Gutierrez</mpeg7:FamilyName>
   </mpeg7:Name></mpeg21:UserInfo>
  <mpeg21:ContentPreferences>
   <mpeg21:UserPreferences>
    <mpeg7:FilteringAndSearchPreferences>
     <mpeg7:CreationPreferences>
      <mpeg7:Creator>
       <mpeg7:Role href="urn:mpeg:mpeg7:cs:RoleCS:2001:DIRECTOR" />
      <mpeg7:Agent xsi:type="mpeg7:PersonType"><mpeg7:Name>
          <mpeg7:GivenName>Steven</mpeg7:GivenName>
          <mpeg7:FamilyName>Spilberg</mpeg7:FamilyName>
       </mpeg7:Name></mpeg7:Agent></mpeg7:Creator>
     </mpeg7:CreationPreferences>
     <mpeg7:ClassificationPreferences>
      <mpeg7:Language>en</mpeg7:Language>
      <mpeg7:Genre href="urn:mpeg:mpeg7:cs:GenreCS:2001:6.7">
       <mpeg7:Name>Romance comedy</mpeg7:Name>
      </mpeg7:Genre>
      <mpeg7:Genre href="urn:mpeg:mpeg7:cs:GenreCS:2001:6.11">
       <mpeg7:Name>Fantasy</mpeg7:Name>
      </mpeg7:Genre>
     </mpeg7:ClassificationPreferences>
    </mpeg7:FilteringAndSearchPreferences>
   </mpeg21:UserPreferences>
  </mpeg21:ContentPreferences>
 </mpeg21:Description>
</mpeg21:DIA>
```

Figure 13. Example of user characteristics description (with user preferences).

due to the lack of access to an operator's MMSC. To perform this simulation we have used many tools of the Nokia Forum[27]: an MMSC emulator, an MMS Java Library, and some terminal emulators.

The MPEG-7 (and MPEG-21) user preference descriptions dealing with classification preferences are used when querying or automatically filtering the content description database. After this selection of content, the MPEG-7 media description preferences are used for adapting the content to the context of use defined via MPEG-21 usage environment descriptions (mainly the terminal characteristics).

As was explained before in Secton 5.3, a final adaptation to the actual characteristics of the terminal is performed before the delivery in the case of MMS service.

7. CONCLUSIONS

We have presented the current features and status of the multimedia messaging subsystem of DTE of the DYMAS system, which is responsible for providing the means for receiving personalized contents on mobile MMS terminals as multimedia messages. The system makes use of the SMIL standard to build summaries of the MPEG-2 contents in the database. These previews are multimedia presentations composed of images and audio clips extracted from these MPEG-2 contents. The system uses MPEG-7 and MPEG-21 standards to personalize the delivery and browsing services according to the preferences of the user and the terminal capabilities.

We detailed the implementation of the main functionalities of the system. These functionalities include registration of new users, terminals, and contents; advanced search; video summary template generation; MMS/SMIL presentation building; personalized delivering of multimedia contents (to e-mail clients and MMS terminal emulators); and browsing of SMIL-based video summaries in the three scenarios: web portal, e-mail, and mobile terminal browsers.

Preliminary testing of the currently implemented and integrated applications (web for fixed PCs,[16] J2ME applications,[17] and multimedia messaging via SMIL and MMS[18]) indicates that the MPEG-2 database can be accessed via alternative networks and services providing content adaptation to terminal characteristics and personalization to user preferences, that is, UMA functionalities. The main problem still remains the requirement for "timely" annotation of content for providing the personalization to user preferences. Work toward semiautomatic real-time annotation will be integrated in the near future. After this, real testing with users and validation of the current UMA functionalities will follow.

Acknowledgments

This work was partially supported by the Ministerio de Ciencia y Tecnología of the Spanish Government under Project TIC2002-03692 (DYMAS). Part of the work of Marta Padilla was partially supported by Amena under a grant of the "Cátedra Amena" of the Universidad Politécnica de Madrid.

References

1. Vetro A, Christopoulos C, Ebrahimi T, guest editors. Universal Multimedia Access (special issue). IEEE Signal Process Mag 2003;20.
2. van Beek P, Smith JR, Ebrahimi T, Suzuki T, Askelof J. Metadata-driven multimedia access. IEEE Signal Process Mag 2003;20:40–52.
3. Martínez JM, Koenen R, Pereira F. MPEG-7: The generic multimedia content description standard. IEEE Multimedia 2002;9:78–87.
4. Burnett I, van de Walle R, Hill K, Bormans J, Pereria F. MPEG 21: Goals and achievements. IEEE Multimedia 2003;10:60–70.
5. DVB home page. Available at: http://www.dvb.org.
6. Mohan R, Smith JR, Li CS. Adapting multimedia internet content for universal access. IEEE Trans Multimedia 1999;1:104–114.
7. Smith JR. Universal multimedia access. In: Proc SPIE Multimedia Systems and Applications IV, vol. 4209; Nov. 2000. pp 21–32.

8. Perkis A, Abdeljaoued Y, Christopolous Ch, Ebrahimi T, Chicharo J. Universal multimedia access from wired and wireless systems. Circuits Syst Signal Process 2001;20:387–402.
9. Fossbakk E, Manzanares P, Yago JL, Perkis A. An MPEG-21 framework for streaming media. In: Proc IEEE Multimedia Signal Processing, October 2001. pp 147–152.
10. Perkis A, Zhang J, Halvorsen T, Kjode JO, Rivas F. A streaming media engine using digital media adaptation. In: Proc IEEE Multimedia Signal Processing, December 2002. pp 73–76.
11. Martínez JM, González C, Fernández O, García C, de Ramón J. Towards universal access to content using MPEG-7. In: Proc ACM Multimedia 2002 Conf; 2002. pp 199–202.
12. Böszörményi L, Hellwagner H, Kosch H, Libsie M, Podlipnig S. Metadata driven adaptation in the ADMITS project. Signal Process Image Commun 2003;18:449–766.
13. Tseng BB, Lin CY, Smith JR. Using MPEG-7 and MPEG-21 for personalizing video. IEEE Multimedia 2004;11:42–52.
14. Magalhaes J, Pereira F. Using MPEG standards for multimedia customization. Signal Process Image Commun 2004;19:437–456.
15. Bescós J, Martínez JM, García N. Real-time audiovisual feature extraction for on-line service provision over DVB streams. Visual Content Processing and Representation (VLBV03), Lecture Notes in Computer Science, vol 2849. Berlin: Springer Verlag; 2003. pp 141–148.
16. Herranz L, Martínez JM. Towards universal and personalized access to audiovisual content in the DYMAS system. In: Proc of 5th Int Workshop on Image Analysis for Multimedia Interactive Services WIAMIS'04 (CD-ROM), Lisbon, April 2004.
17. García A, Martínez JM, Herranz L. Universal and personalized access to content via J2ME terminals in the DYMAS system. Image and Video Retrieval (CIVR04), Lecture Notes in Computer Science, vol 3115. Berlin: Springer Verlag; 2004. pp 225–233.
18. Padilla M. Sistema de transcodificación de vídeos para mensajería multimedia (Video transcoding system for multimedia messaging). Master's thesis, E.T.S.I.T, Universidad Politécnica de Madrid, February 2004.
19. Synchronized multimedia integration language. W3C Recommendation (SMIL 2.0), August 2001. Available at: http.//www.w3.org.
20. MPEG MDS Group. MPEG-7 multimedia content description interface, Part 5: Multimedia description schemes IS. ISO/IEC N4242, July 2001.
21. MPEG MDS Group. MPEG-21 multimedia framework, Part 7: Digital item adaptation FCD. ISO/IEC N5845, July 2003.
22. Westermann U, Klas W. An analysis of XML database solutions for the management of MPEG-7 media descriptions. ACM Comput Surv 2003;35:331–373.
23. Bescós J. Real-time shot change detection over on-line MPEG-2 video. IEEE Trans Circuits Syst Video Technol 2004;14:475–484.
24. Navas CA. Detección de cambios de toma en secuencias de vídeo basado en la banda sonora (Shot boundary detection based on soundtrack). Master's thesis, E.T.S.I.T, Universidad Politécnica de Madrid, May 2002.
25. Cohen A, Kim M, editors. Synchronized multimedia integration language. W3C Recommendation (SMIL 2.0), SMIL 2.0 Basic Profile and Scalability Framework, August 2001.
26. MPEG Requirements Group. MPEG-7 profiles under consideration. ISO/IEC N6039, October 2003.
27. Forum Nokia home page. Available at: http://forum.nokia.com.