Applicability of MPEG-7 Descriptors to Film Heritage

Y. Cobos, M.T. Linaza, A. García, I. Torre
Dept. of Tourism, Heritage and Creativity - Visual Communication Technologies VICOMTech
{ycobos, mtlinaza, agarcia, itorre}@vicomtech.org

Abstract

High-level metadata provide a way to manage, organize and retrieve multimedia data using content descriptions. Within this context, the MPEG-7 Semantic Description Scheme provides tools for storing expressive and interpretable high-level metadata. Despite the flexibility of the MPEG-7 standard, there are very few applications on areas such as Film Heritage. This paper presents the implementation of a user-friendly tool for MPEG-7 based semantic indexing of audio-visual content provided by some European cities with a strong connection with the film sector.

1. Introduction

MPEG-7 is an excellent choice for describing audio-visual content in many applications, mainly because of its comprehensiveness and flexibility. As it has been designed for a broad range of applications and thus, employs very general and widely applicable concepts, the standard comprises a large set of tools for diverse types of annotations on different semantic levels.

The flexibility of the MPEG-7 standard makes it appropriate for many application areas without imposing strict constraints on the metadata models of these sectors. Its flexibility relies strongly on the structuring tools and allows the description to be modular and on different levels of abstraction. MPEG-7 supports fine grained descriptions, allowing the attachment of descriptors to arbitrary segments on any level of detail of the description.

Two main problems arise in the practical use of MPEG-7: complexity and limited interoperability. The complexity results from the use of generic concepts, which allow deep hierarchical structures; high number of different Descriptors (D) and Description Schemes (DS); and their flexible inner structure. As a result, learning MPEG-7 is time-consuming and this may cause hesitance in using this standard in products. Moreover, it is quite difficult to implement tools for working with MPEG-7, and the resulting lack of tools and implementations enhances the hesitance mentioned before.

On the other hand, films are unquestionably part of our heritage. Current systems for accessing such contents related to film objects include the following limitations: distributed sources which store huge amounts of information; different content formats, ranging from traditional ones such as paper to advanced multimedia objects; and finally, and what is more crucial for the content providers, lack of systems which support the user needs such as enriched content, interaction with the information, usability, and exchange of experiences with other users.

Our work analyses the implementation of audio-visual standards in new potential areas such as Film Heritage. Therefore, this paper presents the implementation of a user-friendly tool for MPEG-7 based semantic indexing so that non-expert users from some European cities with a strong connection with the film sector can index their audio-visual content in a simple and accessible way.

The paper is organized as follows. Section 2 summarises some other projects and tools related to this work. The CINeSPACE project is outlined in Section 3 as the application scenario for the implemented tool. Section 4 explains the reasons for selecting MPEG-7 as the standard to perform this task, including the selection of the main descriptors. In Section 5, the CINeSPACE annotation tool is described. Finally, Section 6 presents conclusions and future work.

2. Related work

MPEG-7, "Multimedia Content Description Interface", is an ISO/IEC standard (ISO/IEC JTC1/SC29/WG11) developed by the Moving Picture Experts Group (MPEG) [1]. The standard is used for describing multimedia content data which allows indexing, searching and retrieving information about

the described contents in both real-and non-real time environments. MPEG-7 does not target any application in particular, but supports a broad range of applications.

There are also numerous annotation tools for the creation of MPEG-7 documents available. Within the Intelligent Multimedia Database IMB [2], which focuses on video data, semantic annotation is supported through the integration of parts of Caliph, a "Common and Lightweight Interactive Photo" annotation tool [3] that supports MPEG-7 based structured annotation of images using graph structures. Another representative project is Mecca [4], which supports the annotation of videos based on a growing and evolving ontology for collaborative knowledge creation. Despite dealing with semantic knowledge about video assets, the MPEG-7 descriptors for semantic annotations are not used.

Regarding video annotation based on the MPEG-7 standard, there are plenty of tools such as IBM's MPEG-7 Annotation Tool [5] which focuses on the extraction of high-level metadata, extracting up to 200 different semantic concepts from video streams automatically. The video annotation system MovieTool developed by Ricoh [6] supports hierarchical segmentation within a timeline-based representation of the video. Although the tool is the most mature and complete of the systems, the user interface is too difficult as it is closely tied to the MPEG-7 specifications.

ZGDV's VIDETO [7] has been developed as a research tool to generate video metadata or to test a video server and the retrieval module. It is a video annotation system that hides the complexity of MPEG-7 as the description properties are based on a simple description template mapped to the standard. The LogCreator [8] of the COALA project is a Web based video annotation system which supports video descriptions, providing automatic shot detection and an interface for hierarchical segmentation of videos that can be uploaded to the server, where it is saved as MPEG-7 in a native XML database.

OntoMat-Annotizer [9] is another representative tool developed within the aceMedia project including a user-friendly tool, which allows the semantic annotation of images and videos for multimedia analysis and retrieval, supporting the initialization and linking of RDF(S) domain ontologies with low-level MPEG-7 visual descriptors.

Moreover, Muvino [10] is an MPEG-7 Video Annotation Tool consisting of an embedded video player based on the ViTooKi library and an XML tree view allowing navigation through an MPEG-7 document. A video clip can be divided temporally into a hierarchy of semantic sections, so that the tool helps in the creation and annotation of MPEG-7 VideoSegment descriptions.

Concerning existing applications for the Film and Heritage sector, there are very few examples which have implemented MPEG-7 based annotation techniques. The SIMIPE-Ciné [11] project aims at implementing a Web-enabled software prototype for searching using geographical and visual content criteria over various types of data about the site (annotations, photos, panoramic views, audio clips, and geographical information). The system assists location managers in retrieving more efficiently the sites fulfilling visual and logistical requirements for a particular film shooting.

Finally, ACIS is a system based on an MS Access database that has been designed to support learning communities in the area of Cultural Heritage. The learning process has been disrupted by the civil war and the Taliban regime in Afghanistan [12]. Various multimedia formats such as photos, panoramic views, maps and geographic information, have been described using the MPEG-7 standard.

3. Application scenario: the CINeSPACE project

The CINeSPACE project [13] aims at designing and implementing a mobile rich media collaborative information exchange platform, scalable, accessible through a wide variety of networks, and therefore, interoperable and location-based for the promotion of Film Heritage.

CINeSPACE will enable users to interact with location-based multimedia contents while navigating a city (Figure 1). Audio-visual information will be delivered through a small low-cost wireless Head Mounted Display with a high definition screen situated near-the-eye, and audio phones. CINeSPACE will also comprise a small camera capable of recording and sending what the user is watching. This information can be uploaded to a database through a WLAN hot spot or a 3G connection in order to create collaborative experiences with other users.

The system will address three target users: film tourists who choose a tourist destination due to its relationship with cinema (famous International Film Festivals, film-induced tourism); cinema professionals and film producers who search for possible locations and film facilities in some urban environments; and citizens who have stories to tell, so that they can enhance their community memory.

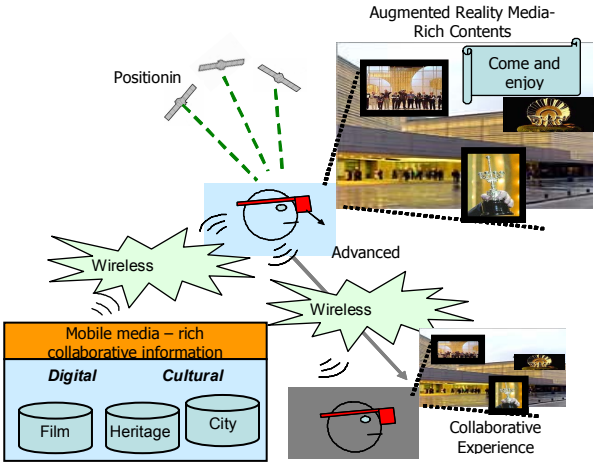

Fig. 1. Schema of the CINeSPACE prototype.

Figure 2 shows the architecture of the indexing and retrieval processes [14]. Multimedia content is stored in a Rich-Media Server in each city (Glasgow, Venice and San Sebastian). Although not focused specifically in optimizing annotation, CINeSPACE will provide the cities with tools to manually annotate the multimedia contents that will be available for the prototype.

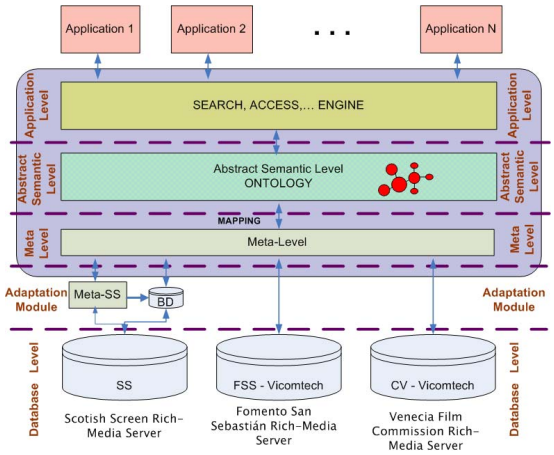

Fig. 2. Architecture of the CINeSPACE Content Management System.

Once the content is indexed with the CINeSPACE annotation tool described in the Section 5 of this paper, the metadata will be used to query the system.

4. CINeSPACE and MPEG-7

4.1. Motivation for using MPEG-7

One of the major challenges of the CINeSPACE project is the design of a suitable data and metadata structure that supports semantic queries using content-based retrieval tools. Moreover, the metadata schema should include mechanisms to link all the material related to a city to its description.

Since an important part of the CINeSPACE metadata concerns the description of audio-visual material (pictures, videos), working within the MPEG-7 framework appears to be a good approach. Thus, we have chosen to explore the usability of the MPEG-7 schema for this application.

In fact, this standard comprises predefined Descriptors and Description Schemes specifically tailored for describing both structural and semantic aspects of images, as requested by CINeSPACE. It also provides constructs for classification schemes that define sets of controlled terms forming vocabularies for particular domains or applications. Furthermore, MPEG-7's DDL offers all the XML advantages such as platform independence and human readability.

4.2. Requirements for the CINeSPACE metadata

Although the location of the user of the CINeSPACE prototype will be the first query filter to retrieve geo-reference content to be delivered within the CINeSPACE system, other aspects of the content retrieval will include content personalization depending on the profile of the user and visual appearance of the content in relation to colour (B&W or colour) and the final visualization device (pocket PC/PDA or binoculars).

Information search results will depend on the exact location of the user of the CINeSPACE prototype (Glasgow, San Sebastian and Venice). For example, when the visitor is in San Sebastian, he/she will be unable to access the information from Glasgow or from Venice. Moreover, some visitors may not want to view a certain image or video more than once, so the usage history of the user must be recorded.

User personalization of the content refers to the type of information that will be provided to the final user depending on his/her profile.

A taxonomy of user profiles will be defined so that content provided to a profile is personalized taking into account some demographical and sociological aspects. Therefore, the personalization parameters will include the preferred language, so that multimedia contents, specially the ones that include audio, will be indexed and only retrieved when the language is selected.

Finally, basic low-level features will be used, as audio-visual contents may be B&W (such as historical film content) or coloured. Moreover, the contents will be available for two types of visualization devices (PDA-type device or binoculars). Therefore, the

CINeSPACE approach should include some additional descriptor to take this issue into account.

4.3. MPEG-7 descriptors for CINeSPACE metadata

According to the MPEG-7 philosophy, all metadata can be represented by means of Descriptors. As many types of metadata can be associated to a given content, there are many types of MPEG-7 Descriptors available. Considering the CINeSPACE project, three types of descriptions can be considered: structural description (structural information given by the multimedia material such as visual descriptors or creation and production information), classification description (characteristic keywords given in a certain domain) and semantic description (set of related semantic entities limited to a certain abstraction level).

This section presents a brief summary of the data and metadata that have been identified for the task of annotating multimedia content in CINeSPACE and the relation with MPEG-7 descriptors.

4.3.1. Camera metadata. Exchangeable Image File (EXIF) and IPTC Information Interchange Model (IIM) metadata record the parameters of the camera at the point a photograph has been taken, including aperture setting, focal length of the lens, exposure time, time of photo, flash information, camera orientation (portrait/landscape) and focal distance. Some of the main descriptors selected for such metadata are *CreationTool* and *CreationCoordinates*. It must be mentioned that these metadata will only be applied to images within the project.

4.3.2. Data from the creation and production processes. These data describe the creation and classification of the multimedia contents, such as director, title, date and location. MPEG-7 provides a tool for structured textual annotation by including specific fields corresponding to questions such as "Who? Where? When? What Action?" that are used for the CINeSPACE metadata.

4.3.3. Global positioning data. GPS data can be recorded live in EXIF, or alternatively GPS track logs can be used to determine the location accurately.

4.3.4. Low-level features. Low-level image analysis aspects such as the colour distribution (*ColorLayout*), colour histograms (*ScalableColor*), edginess (*EdgeHistogram*) or dominant colours (*DominantColor*) can also be used to infer information about the subjects of a picture or a movie shot.

4.3.5. Semantic features. Conceptual information of the reality captured by the content will include the possibility of defining semantic relationships among people, places, objects and interactions between objects. Regarding our scenario, the SemanticBase *DS* includes a number of specialized DSs which describe those specific types of semantic entities (*AgentObject*, *Event*, *SemanticTime, SemanticPlace*).

4.3.6. User preferences. Gaining information about the target users of the content is very important within the CINeSPACE project. Descriptors to describe user preferences and usage history related to the consumption of the multimedia material have been selected from the User interaction Description Tool, mainly the *UserPreference DS*.

5. The CINeSPACE annotation tool

This section presents the extension process of the Caliph tool to fulfil the CINeSPACE requirements in the metadata and associated MPEG-7 Descriptors. Caliph serves as the basis for the work developed within CINeSPACE [3], as it integrates an editor for semantic descriptions based on the MPEG-7 Semantic DS. The editor uses the metaphor of "drawing" a concept graph with semantic objects as nodes and semantic relations as arcs.

The current version of Caliph is restricted to image annotation and this may limit its application to the CINeSPACE project. However, it must be mentioned that the film content will not be segmented to be indexed, but will be treat as an image during the annotation process due to its short duration (no more than 30 seconds). Therefore, the selected MPEG-7 Descriptors will still be applicable, except the camera metadata and the low-level features.

Based on the graphical interface of Caliph, the CINeSPACE annotation tool includes five panels to describe both semantically and structurally the content, namely the "Image Information" panel, which displays the EXIF tags and values and holds the creator of the content; the so-called "Semantics" panel for defining semantic relationships among people, places, objects and interactions between them; the "Shape" and "Visuals" panels for creating *ColorLayout* and *ScalableColor* Descriptors, which are extracted from the image on first loading; and the "User Preferences" panel to describe user preferences pertaining to the consumption of the multimedia material.

Figure 3 shows the graphical interface of the extended "Image Information" panel. Although Caliph is able to automatically extract existing EXIF and

IPTC IIM annotations from images and convert them into valid MPEG-7 Descriptors, this functionality has not been fully exploited due to the historical origin of the content that is used in the CINeSPACE project. Moreover, the *StructuredText* Descriptor has been reduced to four categories (Who, Where, When and What Action), while the possibility of adding *FreeText* Descriptor is available.

The CINeSPACE tool includes two main extensions. On the one hand, new metadata related to the requirements of the project have been added, such as a descriptor for the colour of the multimedia content (B&W, or coloured); a unique private identifier to classify each multimedia content in a MySQL database; and a further Descriptor for the appropriate CINeSPACE device that will render the content.

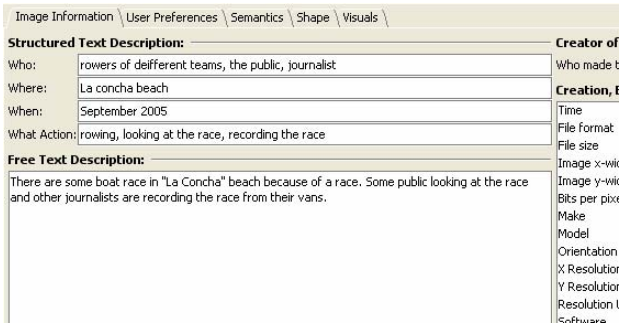

Fig. 3. The "Image Information" panel of the CINeSPACE annotation tool.

On the other hand, as CINeSPACE content has been provided by many institutions, partners and non-partners of the consortium, one of the main problems that arose was related to sharing the content is the digital rights associated to each piece. Therefore, a specific MPEG-7 *ContentUsage* Descriptor has been added to the annotation tool to specify the usage rights.

The core element of Caliph is the so-called "Semantics" panel, which permits defining semantic objects like agents, places, events and times. Semantic objects are used for creating the semantic description by dragging and dropping them onto the panel. This panel has not been extended as the editor fulfils the requirements of the CINeSPACE project.

The "Shape" and "Visual" panels have been limited to a smaller number of descriptors which are extracted automatically by the tool, such as the *EdgeHistogram* (edginess), the *DominantColor* (dominant colours), the *ColorLayout* (colour distribution in an image) or the *ScalableColor* (basically a colour histogram), as the CINeSPACE project does not focus on low-level descriptors. Finally, a new "User Preferences" panel dedicated to user preferences metadata has been added to the Caliph basic editor (Figure 4).

From the *UserPreferences DS*, we have selected the *CreationPreferences DS* to specify user preferences related to the *Period* (when the content was created); *Keywords* or Georeference data (longitude and latitude) and the *ClassificationPreferences DS* to define the *Genre* (comedy, music, news) and the *Subject*. The defined metadata will allow information retrieval based mainly on the geographical location of the CINeSPACE device user and his/her interests.

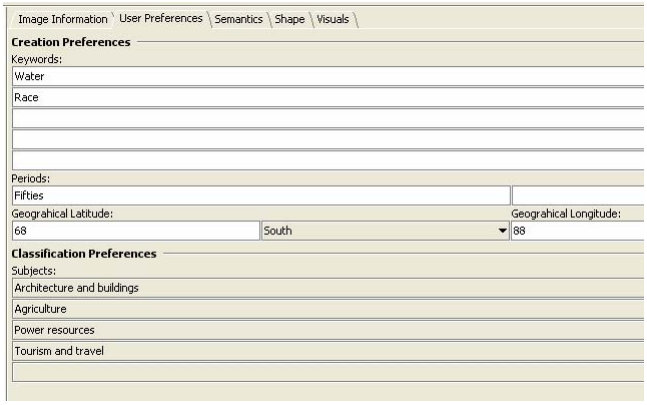

Fig. 4. The "User Preferences" panel of the CINeSPACE annotation tool.

Although Caliph supports pre-annotation of sets of images using the so-called autopilot, the annotation process will be manually carried out within CINeSPACE. It is very important that the cities providing the content have overall control over the Descriptors in order to efficiently retrieve the content. Once the multimedia content has been fully described, each annotation of the multimedia content can be saved as an MPEG-7 XML file.

6. Conclusions and future work

This paper describes the implementation of the CINeSPACE annotation tool which enables semantic and structural annotations of multimedia content for Film Heritage. Since an important part of the CINeSPACE metadata concerns the description of audio-visual material, the MPEG-7 standard has been selected as the development framework. Three types of descriptions have been considered: structural description, classification description and semantic description.

After analyzing the existing approaches for multimedia annotation, the open source tool called Caliph has been selected as the basis for the work developed within the project. The annotation process will be manually carried out within CINeSPACE. It is very important that the cities providing the content

have overall control over the Descriptors in order to efficiently retrieve the content.

Regarding the annotation tool, several extensions have been added to the basic Caliph software. Firstly, new metadata related to the requirements of the project have been added, such as a Descriptor for the colour; a unique private identifier to classify each multimedia content in a MySQL database; and a further Descriptor for the appropriate CINeSPACE device that will display the content. Moreover, a specific MPEG-7 Descriptor named Content Usage has been added to the annotation tool in order to take into account the rights of the content owners. Finally, a new panel dedicated to user preferences metadata has been added to the Caliph basic editor.

Concerning future work, the CINeSPACE annotation tool has been distributed and tested among the final users of the project, analysing the viability of such tools in order to annotate multimedia content databases. Although the tool has not been widely used by the non-experts from the cities, first feedbacks are quite positive towards the new tool. Moreover, the tool will be extended to be used in other sectors dealing with large multimedia digital libraries such as the tourism or cultural heritage sectors.

7. Acknowledgements

This work has been partially funded under the 6th Framework Programme of the European Union within the IST project CINeSPACE (IST FP6- 034990). Authors would also like to thank the Basque Government for the funding of the eTourgune project. This work has been done within the COST Action 292.

8. References

[1] ISO/IEC JTC1/SC29/WG11 Coding of Moving Pictures and Audio, http://www.chiariglione.org/mpeg/standards/mpeg-7/mepg-7.htm.

[2] W. Klieber, K. Tochtermann, M. Lux, H. Mayer, H. Neuschmied, and W. Haas, “IMB - Ein XML-basiertes Retrieval Framework für digitales Audio und Video“, Berliner XML Tage 2003, Berlin, Germany.

[3] M. Lux, W. Klieber, and M. Granitzer, “On the Complexity of Annotation with the High Level Metadata”, *Journal of Universal Knowledge Management*, 1(1), 2006, pp. 54-58.

[4] R. Klamma, M. Spaniol, and M. Jarke, “MECCA: Hypermedia Capturing of Collaborative Scientific Discourses about Movies”, *The International Journal of an Emerging Discipline* 8, 2005, pp. 3-38.

[5] IBM MPEG-7 Annotation Tool. http://www.alphaworks.ibm.com/tech/videoannex.

[6] Ricoh MovieTool. http://www.ricoh.co.jp/src/multimedia/MovieTool/about/index.html#outline.

[7] Zentrum fuer Graphische Datenverarbeitung, VIDETO-Video Description Tool. http://www.zgdv.de/zgdv/zgdv/departments/zr4/Produkte/videto.

[8] Swiss Federal Institute of Technology, COALA (Content-Oriented Audiovisual Library Access)–LogCreator. http://ditwww.epfl.ch/SIC/SA/publications/FI03/fi-4-3/4-3-page8.pdf.

[9] N. Carsten, S. Timmerman, D. Kosmas, and Y. Anastasopoulos, “M-Onto-MatAnnotizer: Linking Ontologies with Multimedia Low-Level Features for Automatic Image Annotation”, *Proceedings of 3rd European Semantic Web Conference (ESWC 2006)*, Budva (Montenegro), June 2006, pp. 633-640.

[10] MuViNo – an MPEG-7 video annotation tool, part of the ViTooKi open-source project, September 2004, http://vitooki.sourceforge.net/components/muvino.

[11] V. Gouaillier, L. Gagnon, S. Paquette, and P. Poullaouec-Gonidec, "Use of the MPEG-7 Standard as Metadata Framework for a Location Scouting System - An Evaluation Study", *Proceedings of the International Conference on Dublin Core and Metadata Applications (DC-2005)*, Madrid (Spain), September 2005.

[12] R. Klamma, M. Spaniol, M. Jarke, Y. Cao, M. Jansen, and G. Toubekis, “A Hypermedia Afghan Sites and Monuments Database”, *Lecture Notes in Geoinformation and Cartography*, Springer Berlin Heidelberg, Location, 2006, pp. 189-209.

[13] CINeSPACE Project http://www.cinespace.eu.

[14] P. Santos, A. Stork, M.T. Linaza, O. Machui, D. McIntyre, and E. Jorge, “CINeSPACE: Interactive Access to Cultural Heritage while on-the move”, *Proceedings of the HCI International 2007*, Beijing (China), July 2007.