

An Intelligent System for Retrieval and Mining of Audiovisual Material Based on the MPEG-7 Description Schemes

Giorgos Akrivas, Spiros Ioannou, Elias Karakoulakis, Kostas Karpouzis, Yannis Avrithis,
Anastasios Delopoulos, Stefanos Kollias, Iraklis Varlamis and Michalis Vaziriannis
Department of Electrical and Computer Engineering
National Technical University of Athens
9, Iroon Polytechniou Str., 157 73 Zographou, Athens, GREECE
email: sivann@image.ece.ntua.gr

ABSTRACT: A system for digitization, storage and retrieval of audiovisual information and its associated data (metainfo) is presented. The principles of the evolving MPEG-7 standard have been adopted for the creation of the data model used by the system, permitting efficient separation of database design, content description, business logic and presentation of query results. XML Schema is used in defining the data model and XML in describing audiovisual content. Issues regarding problems that emerged during system design and their solutions are discussed, such as customization, deviations from the standard MPEG-7 DSs or even the design of entirely custom DSs. Although the system includes modules for digitization, annotation, archiving and intelligent data mining, the paper mainly focuses on the use of MPEG-7 as the information model.

KEYWORDS: Audiovisual archives, multimedia databases, multimedia description schemes, MPEG-7, retrieval and mining of audiovisual data.

INTRODUCTION

Current multimedia databases contain a wealth of information in the form of audiovisual and text data. Even though efficient search algorithms have been developed for either media, the need for abstract data presentation and summarization still exists [1]. Moreover, retrieval systems should be capable of providing the user with additional information related to the specific subject of the query, as well as suggest other, possibly interesting topics. The MPEG-7 standard [2] aims to satisfy the above operational requirements by defining a multimedia content description interface and providing a rich set of standardized tools to describe multimedia content. Unlike previous MPEG standards (MPEG-1/2/4), MPEG-7 descriptors do not depend on the ways the described content is coded or stored; it is even possible to create an MPEG-7 description of an analogue movie or of a picture that is printed on paper [3]. Moreover, automatic or semi-automatic feature extraction algorithms will be outside the scope of the standard, similarly to previous MPEG standards. For some features, such as textual description, human intervention seems unavoidable for the foreseeable future.

MPEG-7 will specify a standard set of Descriptors (Ds) that can be used to describe various types of multimedia information. It will also specify a rich set of predefined structures of Descriptors and their relationships, as well as ways to define one's own structures; these structures are called Description Schemes (DSs). Defining new Description Schemes is done using a special language, the Description Definition Language (DDL), which is also a part of the standard. At the 51st MPEG meeting in Noordwijkerhout, it was decided to adopt the XML Schema Language [4] with a set of minimal MPEG-7-specific extensions as the MPEG-7 DDL [5]. The standard also defines a set of DSs and Ds, which every MPEG-7 parser should read.

In this paper we present a system for efficient digitization, annotation, storage, search and mining of audiovisual data from large distributed multimedia databases through several types of networks. The system has been developed in the context of a Greek project named PANORAMA. The main objectives of the project were the interoperability of databases and the availability of software products and services on networks for open multimedia access. Digitization of audiovisual archive assets provides information of all possible media (video, images, sound, texts, etc.) for wide public and professional use.

In order to deal with design, description and management issues, as well as navigation and retrieval in multimedia entities, we have adopted concepts included in MPEG-7, such as Multimedia Description Schemes (MDS). The

integrated architecture we have developed can thus be smoothly integrated in MPEG-7 compatible multimedia database systems. Moreover, the adoption of MPEG-7 description schemes in all system modules permits efficient separation of database design, content description, business logic and presentation of results without having to rearrange the employed schemes. In this paper we present the issues regarding the use of MPEG-7 as the information model.

DESCRIPTION SCHEME DESIGN

The status of MPEG-7 had provided a sufficient information model, as well as DDL, a language for creating new data types. Using DDL, MPEG-7 had defined a set of Descriptors for describing features of audiovisual content, and Description Schemes for organizing hierarchical description information. However, specific needs of the PANORAMA users, as well as the incompleteness due to the non-final stage of the standard, had forced us to modify most of the DSs and create a number of new ones.

The most basic Description Schemes used by PANORAMA are described below:

A description of a whole Audiovisual document is contained in the AudiovisualDS. An Audiovisual DS contains information about physical storage (MediaInfoDS), textual information (MetaInfoDS), structural information (SyntacticDS) and summarization information (SummarizationDS).

Regarding the media, we had to deal with the following situations:

- A medium (e.g. a film, cassette, et.c.) can contain many audiovisual entities
- An audiovisual entity can be segmented in many different media
- A medium can consist of many medium segments (e.g. a film consisted of many film pieces)
- Several copies of the same medium can exist

Because of the above complexity, (especially in old cinema films), we had to store information about physical mediums in a separate entity, MediumDS, unlike MPEG-7. Specializations of this type have been CassetteDS, FilmDS and PhotoDS, as well as the two digital types, DigitalVideoDS and DigitalImageDS. Each of the media can contain one or more MediumSegmentDS, which has its own frame numbering. Copies between pairs of mediums are stored in MediaRelationDS. Information about where audiovisual data is recorded is stored in MediaInfoDS. Copies of the audiovisual data in various mediums are stored in a separated MediaProfileDS, which contains pointers to MediumDS. It can also contain frame numbers to specific MediumSegmentDS.

Information about segmentation of the information into separate units, such as shots, is stored in SyntacticDS. Unlike MPEG-7, which uses an unlimited number of levels, PANORAMA uses by specification a four-level hierarchy of themes, shot groups, shots and keyframes. Each of the units can contain its own MetaInfo and MediaInfo.

Summarization information is contained in SummarizationDS. Like MPEG-7, a sequential and a hierarchical summarization are supported. The former contains pointers to a selection of shots and keyframes, while the latter consists of a subtree of the whole segmentation tree.

Annotation information desired by the users of the system includes strings (StringValueDS), integer numbers (IntegerValueDS), Dates (DateValueDS), frame numbers (TimeValueDS), boolean values (BooleanValueDS) and enumerations (DynamicListValueDS). Each of these value contains a reference to each name (which is contained in a set of names) and one or more values. These values are stored in a MetaInfoDS.

INTERACTION BETWEEN SUBSYSTEMS

A basic design principle of PANORAMA is that data exchange among subsystems is in the form of valid XML documents. The reasons leading to this decision is that a single parser was sufficient for every data exchange among modules. Use of XML created the need for a language capable of defining valid XML documents, a system for converting XML documents into C++ classes suitable for internal processing and a database capable of storing and retrieving XML documents.

The need for a less narrow-focused XML definition language than DDL, led us to adopt the emerging XML Schema language. XML Schema is more expressive for describing hierarchy and inheritance, which made it a more appropriate choice for defining generic, non-MPEG7 XML documents.

Conversion to C++ classes had been necessary because of the difficulties in using the existing interfaces for handling XML documents in memory, such as the DOM and SAX interfaces. The module that was constructed creates a class for each complex type and a vector for each sequence of elements. Moreover, XML references are converted to C pointers.

The hierarchy and inheritance that is employed by XML Schema creates the need for construction of a new database that extends the currently used relational model, which supports neither of the features. The module that was constructed creates a table of a relational database for each complex type and each of its elements. Association between complex types, inherited complex types and elements is performed via foreign keys. The database supports the COM interface, which makes it accessible by both Windows based and web based applications.

THE ARCHIVING SUBSYSTEM

The archiving subsystem enables the administrator of an audio-visual archive to enter, modify and delete physical mediums (in digital or in analog form). The production of copies is automated by issuing copy orders, where the necessary data (the mediums involved and the time codes) are taken from the database and written into an XML file, which is used by the machine performing the copy. Alternatively, copy can be done manually; in this case, the user enters the information to the system a posteriori.

Using the archiving subsystem, the administrator can also issue annotation orders, to accept annotation results and to cancel pending annotation orders.

Finally, standard database operations (creation, destruction, backup and restore) are supported, as well as directly giving commands (in XML) to the XML database.

USER INTERFACE

The expert person that provides the verbal information about the content (the annotator) is supported by a special application developed for the system. The application takes as input an XML description of the audiovisual content and gives as output the same description enriched by the annotator. The basic features of the annotation application, shown in Figure 7, are the following:

- Automatic shot and keyframe detection
- Tree representation of the video segments (AV entity/Theme/ShotGroup/Shot/Keyframe, summarization and physical storage)
- Usage of the MPEG-1 format for the annotation
- Customisable annotation of each video segment
- Automatic extraction of several low-level features, such as camera motion, moving regions, faces and text



Figure 1: User interface of the annotation tool.

Following the “Whatever is known and needed is included in the description”, the annotator is free to choose the depth in which he decides to annotate. For example, he might annotate only the full AV entity, or annotate part or all of the shots. Or he might not take the time to execute the automatic shot extraction feature and limit himself to the annotation only.

OTHER SUBSYSTEMS

A web – based user interface was constructed. The interface supports user authentication, queries (both text – based and content – based) and presentation of results. The hierarchical nature of the description is mapped in a hierarchy of interlinked web pages. Streaming and downloading is supported for viewing MPEG files.

A digitization system was utilized to convert video cassettes into MPEG video. The digitization subsystem is capable of receiving XML information regarding timecodes of the cassettes to be digitized. Cinema film was converted into TV signal via telecine. Protection of intellectual property in digitized images and video is achieved through a watermarking application that was constructed for PANORAMA. The application embodies a watermark directly into the DCT information of the MPEG and JPEG data. A set of algorithms for automatic extraction of visual features (color, and texture information, moving objects, camera motion, faces and regions with text) was constructed in order to support content – based queries. Moreover, the shots are automatically extracted in order to save the annotator time. These algorithms are used by the annotation subsystem. The XML database handles visual features separately, in order to achieve improved performance for feature – based queries.

EXTENSIONS UNDER CONSIDERATION

This project has focused into structural (syntactic) aspects of multimedia descriptions. While these aspects are suitable for automatic extraction by machines (for example, shot and face detection), they are not oriented towards human requirements. MPEG-7 includes a semantic part of the description. A semantic description is more human oriented, as it deals with entities such as objects, events, and their relations. Moreover, a semantic description is more suitable because it provides the means to map the vagueness that is present in human – produced descriptions. More specifically, the MPEG-7 semantic relations are fuzzy, since two semantic entities can be related to a certain extent. While a semantic query is more desired by the human user, a semantic description is more difficult to achieve automatically. In this domain, a set of fuzzy rules has been proposed; fuzzy rules would generate fuzzy relations out of existing ones.

An innovative system for handling audiovisual information and its associated metadata was presented. It illustrated the potential of the MPEG-7 evolving standard, as a means to formalize the description of audiovisual content. In particular, the use of MPEG-7 descriptions as a means of communications between the system modules proved to be a powerful and efficient tool. The personalization of the user queries as well as the semantic unification of individual archives with custom description schemes using novel neurofuzzy platforms are the objective of other R&D projects that are currently under development at the Image Video and Multimedia systems Laboratory.

REFERENCES:

- [1] Chiariglione L., “MPEG and Multimedia Communications,” IEEE Trans. Circuits and Systems for Video Technology, Vol. 7, Feb. 1997, pp. 5-18.
- [2] ISO/IEC JTC1/SC29/WG11, “MPEG-7 Overview (v. 1.0),” Doc. N3158, Dec. 1999.
- [3] ISO/IEC JTC1/SC29/WG11, “MPEG-7: Context, Objectives and Technical Roadmap, (v.12),” Doc. N2861, July 1999.
- [4] “XML Schema Part 0: Primer,” W3C Working Draft, Sept. 2000 (<http://www.w3.org/TR/xmlschema-0>)
- [5] ISO/IEC JTC1/SC29/WG11, “Text of ISO/IEC CD 15938-2 Information technology – Multimedia content description interface – Part 2: Description definition language,” Doc. N3702, Oct. 2000.
- [6] ISO/IEC JTC1/SC29/WG11, “Text of ISO/IEC 15938-5/CD Information Technology – Multimedia Content Description Interface – Part 5: Multimedia Description Schemes,” Doc. N3705, Oct. 2000.
- [7] G. Votsis, A. Drosopoulos, G. Akrivas, V. Tzouvaras and Y. Xirouhakis, “An MPEG-7 Compliant Integrated System for Video Archiving, Characterization and Retrieval”, IASTED International Conference on Signal and Image Processing (SIP2000), Las Vegas, Nevada, November 2000.