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 (MMDS). 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 NMPEG-7 as the information model.

DESCRIPTION SCHEME DESIGN

Even though the Descriptors (Ds) and Description Schemes (DSs) proposed by the MPEG Group are more than enough for the most systems, they can be extended to suit specific requirements or match existing data and applications. In most cases our design was based on the standard DSs [6] with certain extensions to impose additional constraints or add extra functionality. The hierarchical structure of the system is shown in Figures 1, 2 and 3 in UML format; this format is used here instead of the usual text-based DDL or XML Schema so as to illustrate the employed hierarchy and DSs in a graphical way. The diamond symbol in these figures represents the composition relationship. The range associated to each element represents frequency in the composition relationship, while the arrows denote the class inheritance relationship.
In general, the AudioVisualDS shown in Figure 1, represents an AV entity, i.e. a movie or a picture. AudioVisualDS contains whatever data is known about an existing AV entity. Consequently, all the DSs that it contains are optional: syntactic structure (SyntacticDS), semantics (SemanticDS), links between segments and semantic entities (SyntacticSemanticLinkDS), physical storage (MediaInfoDS), and verbal description (MetaInfoDS).

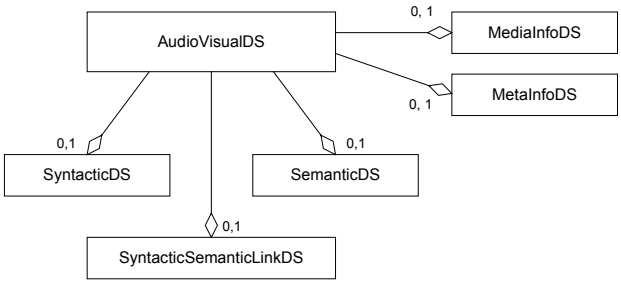

Figure 1: Representation of the AudioVisualDS hierarchy.

AudioVisualDS is designed as a metaphor of the typical method of organizing the content in a written document, i.e. with the use of a Table of Contents and an Index. In such a context, the Table of Contents (syntactic information) aims to define the structure of the archive, as it does in a book or document, using linear syntax regardless of the internal organization of the material and the linking which occurs with respect to its semantic content. On the other hand, the goal of the Index (semantic information) is not to describe the structure of the content but to provide useful references to the actual material. These references are usually not complete, in the sense that the Table of Contents essentially provides access to every piece of information in the archive, but they are selected based on their semantic value to humans and may be recurring for the same item.

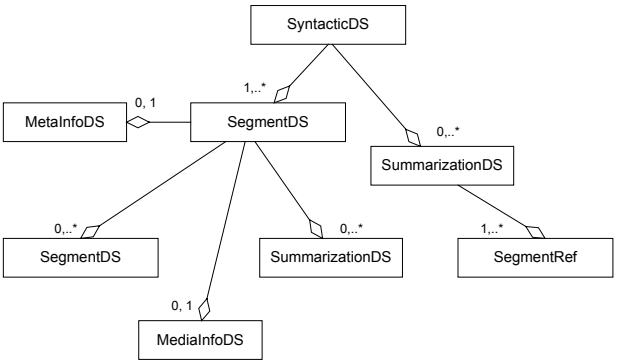

Figure 2: Structure of the audiovisual material in SyntacticDS.

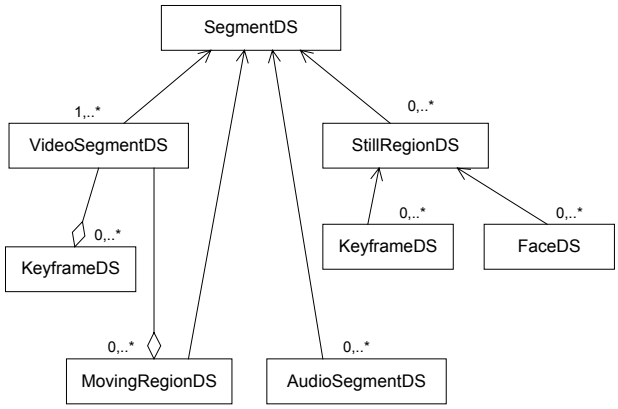

Figure 3: Definition of segment types, through the inheritance mechanism.

In our implementation, syntactic information is contained in the SyntacticDS, shown in Figure 2. The SyntacticDS contains information about the organization of the content in the physical level, as well as signal-based descriptors, such as camera movement or definition of shot groups. The inclusion of recurring SegmentDSs allows the creation of hierarchical TOCs, where the actual material and accompanying meta-information are presented in a way that preserves the required level of abstraction. In essence, the temporal structure
and overall visual properties of a high-level object, i.e. a theme, are represented as a single node and may be decomposed to shot groups or shorter lower-level shots.

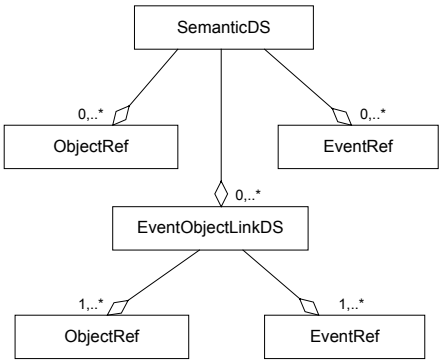

Figure 4: Description of semantic content through a hierarchy of objects and events in SemanticDS.

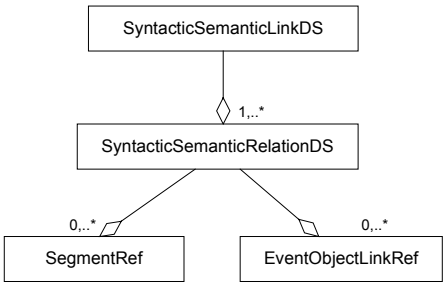

Figure 5: Linking between syntactic and semantic information.

A SegmentDS represents a part of the content that can be thought of as an entity, and therefore it contains a verbal description (MetaInfoDS), a number of other segments, which consist a further segmentation of the content, media information, if needed, and possibly a number of summaries as well. Several kinds of segments are defined through the inheritance mechanism, as shown in Figure 3. These are: VideoSegmentDS, which represents a segment in time (e.g. a shot), StillRegionDS, which represents a still object, MovingRegionDS, which represents a moving object, and AudioSegmentDS, which represents an audible segment. A VideoSegmentDS can contain other VideoSegmentDSs, MovingRegionDSs and KeyframeDSs. Two specializations of StillRegionDS are defined: KeyFrameDS and FaceDS.

Each SegmentDS also contains several low-level (usually automatically extracted) descriptors such as color features, camera motion, shape and texture of moving regions, human face or text areas et.c.
The semantic content is described through the SemanticDS, shown in Figure 4, which contains an hierarchy of objects (ObjectDS) and events (EventDS). The relationship between objects and events is somewhat similar to the relationship between nouns and verbs in natural language. Unlike previous DSs, SemanticDS contains pointers (references) to objects and events, instead of full instances of objects. Therefore, many AV entities can share the same objects (e.g. persons or locations). A SemanticDS can also contain a number of EventObjectLinkDS, which provide a means to link events and objects. Another link is made to semantic entities, the SyntacticSemanticLinkDS shown in Figure 5. This provides a means to relate high-level descriptions (eg a person) to low-level ones (eg a moving region). A SyntacticSemanticLink contains all the SyntacticSemanticRelationDSs, which link segments with event-object links.

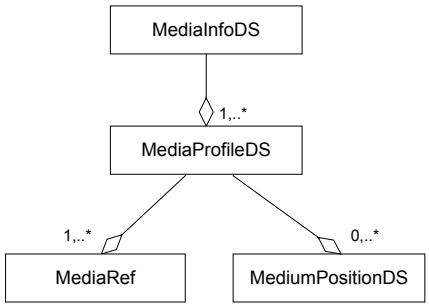

Figure 6: Media storage information and profiles in MediaInfoDS.

CUSTOMIZATION ISSUES

The description schemes discussed so far are mainly based on the standard DSs of the MPEG-7 [6]. However, the description of specific information requires either deviations from the standard DSs or even the design of entirely custom DSs. The main reason for customization is the existence of heavy constraints imposed by the target user of the system (ERT) and the lack of certain information from the standard MPEG-7 DSs at the time of the PANORAMA architectural design. A typical example is the meta-information of audiovisual programs, shown in Figure 6. MetaInfoDS in our system contains all verbal information that is known about an AV entity or a segment. This includes strings (for example names of actors), times (for example time of production), and numbers (for example number of episodes). Each of the variables can have an unlimited number of values, and new variables can dynamically be added. However, information such as program type, genre etc. is given as predefined enumerations customized for the system user.
Another example is the MediaInfoDS which contains information about physical storage, namely the various copies of the content (MediaProfileDS), and the media that these copies are made on (MediaProfileSegmentDS). MediaInfoDS was extended customized to support a wide range of analog and digital media such as films, VHS tapes, digital Betacam, MPEG files, photographs and documents, as well as a large number of format and encoding parameters. Additional restrictions were imposed on the syntactic structure of AV documents by generating a hierarchy of themes, shot groups and shots. Three types of VideoSegmentDS are defined using inheritance, namely, ThemeDS, ShotGroupDS and ShotDS. The SyntacticDS shown in Figure 2 is then composed of ThemeDSs. ThemeDSs optionally contain a number of ShotGroupDSs, which in turn contain ShotDSs. Finally, sequential and hierarchical summarization information is supported by embedding a SummarizationDS in SyntacticDS as well as in each individual SegmentDS (optionally), as shown in Figure 2. The SummarizationDS contains a number of references to objects that can be considered sufficient to represent the full set of segments.

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
- Customizable annotation of each video segment
- Automatic extraction of several low-level features, such as camera motion, moving regions, faces and text

Figure 7: 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.

CONCLUSION

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 user queries as well as the semantic unification of individual archives with custom description schemes are open issues for future research.

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.