Unified Intelligent Access to Heterogeneous Audiovisual Content

A. Delopoulos, S. Kollias, Y. Avrithis

Image, Video and Multimedia Systems Laboratory
National Technical University of Athens Greece
9 Heroon Polytechniou St., 157 73 Zographou,
Greece
Tel: +301-772 3040, Fax: +301-772 2492

E-mail: adelo@image.ntua.gr

W.Haas, K. Majcen

Institute of Information Systems & Information
Management, JOANNEUM RESEARCH
Steyrergasse 17, A-8010 Graz,
Austria
Tel +43 316 876 1119

E-mail:
{WernerHhaas,Kurt.Majcen}@joanneum.at

Abstract. Content-based audiovisual data retrieval utilizing new emerging related standards such as MPEG-7 will yield ineffective results, unless major focus is given to the semantic information level. Mapping of low level, sub-symbolic descriptors of a/v archives to high level symbolic ones is in general difficult, even impossible with the current state of technology. It can, however, be tackled when dealing with specific application domains. It seems that the extraction of semantic information from a/v and text related data is tractable taking into account the nature of useful queries that users may issue. And the context determined by user profile. The European IST project FAETHON is developing a novel platform, that intends to exploit the aforementioned ideas in order to offer user friendly, highly informative access to distributed audiovisual archives.

1 Introduction

It becomes clear among the research community dealing with content-based audiovisual data retrieval and new emerging related standards such as MPEG-7, that the results to be obtained will be ineffective, unless major focus is given to the semantic information level, defining what most users desire to retrieve. Mapping, however, low level, sub-symbolic descriptors of a/v archives to high level symbolic ones is in general difficult, even impossible with the current state of technology. It can, however, be tackled when dealing with specific application domains. It seems that the extraction of semantic information from a/v and text related data is tractable taking into account:

a) The nature of useful queries that users may issue. This is only a portion of the general set of questions related to "content understanding". Using all types of multimedia information of the archives makes the task more tractable.

b) The context determined by user profile.

FAETHON, a European consortium consisting of Sysware S.A. (GR), Lambrakis Research Foundation (GR), Joanneum Research (A), Film Archive Austria (A), Starlab Research NV/SA (B), Oracle GmbH (A), Institute of Communications and Computers Systems National Tech. Univ. of Athens (GR) and the Hellenic Broadcasting Corp. (GR) is developing a novel platform, in the framework of IST projects, that intends to exploit the aforementioned ideas in order to offer user friendly, highly informative access to distributed audiovisual archives.

Digital archiving of multimedia content including video, audio, still images and various types of documents has been recognized by content holding organizations as a mature choice for the preservation, preview and partial distribution of their assets. The advances in computers and data networks along with the success of standardization efforts of MPEG and JPEG boosted the movement of the archives towards the conversion of their fragile and manually indexed material to digital, computer accessible data. By the end of last century the question was not on whether digital archives are technically and economically viable, rather on how digital archives would be **efficient** and **informative.** In this framework, different scientific fields such as, on the one hand, development of data base management systems, and on the other hand, processing and analysis of multimedia data, as well as artificial and computational intelligence methods, have observed a close cooperation with each other during the last few years. The attempt has been to develop intelligent and efficient human computer interaction systems, enabling the user to access vast amounts of heterogeneous information, stored in different sites and archives.

Data base management systems (DBMS) have been designed that are able to handle such types of access to the stored information. Attaching information bits, called metadata, to the original data is the means for achieving this goal. The focus of technological attempts has been on the analysis of digital video, due to its large amounts of spatio-temporal interrelations, which turn it into the most demanding and complex data structure. Current and evolving international standardization activities, such as of the EBU, MPEG-4, MPEG-7, or JPEG-2000 for still images, deal with aspects related to data structures and metadata. In particular, the new MPEG standards are object-oriented, i.e., adopt video objects as the information unit, which is different from the information units used in the current form of video and film, i.e. scenes or shots. Of major importance is the contribution of MPEG-7 and JPEG-2000 to using metadata related to the visual and acoustic content of archived objects.

In more detail, MPEG-7 will define a standard for describing multimedia content. The objective is to quickly and efficiently search and retrieve audiovisual material. To allow interoperability, the standard adopts some normative elements, such as Descriptors (D's), Description Schemes (DS's), the Description Definition Language (DDL) as well as Coding and System Tools. The Descriptors define the syntax and the semantics of the representation of features, while the Description Schemes specify the structure and semantics of the relationships between Descriptors or other Descriptions. Many descriptors have been submitted for MPEG-7, some of which either accepted and included in the eXperimental Model (XM), which is a platform and tool set to evaluate and improve the tools of MPEG-7, or are in the experimentation (Core Experiments, CE) phase. Two parallel levels of descriptors are defined: the syntactic one, which describes the perceptual properties of the content, such as color and motion of spatio-temporal segments and the semantic one, which describes the meaning of content, in terms of semantic objects and events.

Syntactic description seems to be well in hand in MPEG-7, but fleshing out the semantic description has not yet received the required attention. Mapping low-level features, such as color, texture, shape, layout and motion, in the visual data case, or pitch, energy level in the audio data case, with high-level semantic concepts, the latter defining what most users desire to retrieve, is, however, rather difficult, if not impossible. Thus, content-based image retrieval commonly does not produce satisfactory results.

Various European projects have provided audiovisual archiving systems in the philosophy of MPEG-7, based on databases with proprietary documentation and user access interfaces, handling metadata information. Among the first approaches are those of VICAR and DiVan projects. The AVIR project, in turn, has proposed a language for expressing metadata information and description schemes, following up the developments towards the MPEG-7 standard. The AVIR architecture provides a basis for managing and interchanging MPEG-7 descriptions between the content provider, the service provider and the consumer system. A Table of Content plus an Analytical Index are adopted for content description of a/v material. The ACTS DICEMAN project has been developing an MPEG-7 Database implementation evaluating the use of agents in managing information overload.

2 Key Issues in Handling Large and Distributed A/V Archives

2.1 Customization versus Unification

More and more a/v archives are adopting computer based systems to organize their catalogues and the various types of metadata associated to their content; the percentage of the a/v content available in digital form increases as well normative description structures, defined mainly by MPEG-7 and by the recommendations of bodies like EBU, seem to offer a solid common ground for the organization of the archives. By intention, though, the targeted standardization level will provide tools, best practice examples and prototypes but will not prescribe the exact structure, the level of documentation detail and the exact classification schemes of individual archives. It is at this point that developers, archivists and even administration of each archive enter the game and *customize the norms of the standards.* The final choices will reflect (i) the actual nature of the content (e.g., film archives containing mainly fiction features will have different structure compared to television news archives or archives of documentary productions), (ii) the admitted cost of digitization and documentation (e.g., documentation up to the level of video shots may be impossible or even useless with respect to the available resources), (iii) the individual thematic categories/specialization of individual archives, (iv) the status of archiving/digitization progress, (v) the existence or lack and the type of analog media (films, video tapes, paper hardcopies, slides, etc) within the same archive.

Hence, on the one hand, the ability of customization is a virtue of standards like MPEG-7 because otherwise they could reduce to monolithic and finally useless ones. On the other hand unification is a demand from users' side. It is desired that multiple a/v archives within the same content holding organization or across such organizations at a regional, national or European level provide access to their users or clients in a uniform and transparent way. Mining of a/v content, in terms of both searching and retrieval, should be possible irrespectively of the actual owner of the a/v assets avoiding multiple searches and/or reformulation of the queries for each individual archive. It seems that the ability of individual archives to join groups of possibly heterogeneous counterparts will be critical for their viability and their competitiveness within the growing market of digital a/v offer.

2.2 Overwhelming versus Personalized filtering

Future a/v archive users will very soon face the “Internet Syndrome”, that is, when expressing a query they will get thousands of matches as a response. This is essentially a result of the increase in detailed metadata information that will accompany the a/v content. The situation will be even worse if multiple archives are simultaneously accessed by the same query. The **syntactic** information (D's or DS's in MPEG-7 terminology) that will be used to characterize the content will not be useful in expressing more descriptive and narrow queries unless it maps onto semantic descriptions. It is also understandable that this mapping may depend on the context of the query that greatly depends on the **personal profile** of the user. As an example one may think of a low level syntactic description of the form “an orange circle moving downwards”; for a sportscaster this corresponds to a shot of a basketball game while for a producer it may correspond to a segment of a sunset footage.

In addition to the interpretation of low level features, personal profiling may define thematic preferences that determine the priorities of the returned responses. Searching, e.g., with the keyword “accident” may return the available material related both to “car accidents” and to “nuclear accidents”. The priority between them may be implied by the profile of the user. It is interesting to mention that in many cases literal refinement of the query by adding the additional specification “car” or “nuclear” respectively may abandon both responses if those strings were not included in the subject but are rather implied by other descriptors. MPEG-7 foreseeing the need of semantic description of the content, supports definition of DS’s of semantic nature. It seems, though, that at a **normative level** only a set of generic such descriptors will be standardized. Personalization itself is beyond the scope of the standard.

3 Technical Approach

FAETHON will create a novel system exploiting the advances in handling audiovisual (a/v) content and related metadata, as introduced by MPEG-4 and MPEG-7, to offer advanced access services characterized by the tri-fold "semantic phrasing of the request (query)", "unified handling" and "personalized response".

The proposed system will play the role of an intermediate access server residing between end users and multiple heterogeneous archives organized according to new MPEG standards. The core technological target of the project is to blend the achievements in characterizing a/v content - especially visual and acoustical - with innovative hybrid intelligence technologies in order to:

a) offer unified semantic views to existing a/v archives, beyond the classification schemes and subject indexes of each archive,

b) personalize those views according to the retained profile of individual users or specific user groups; the latter clearly appreciating that semantic interpretation heavily relies on the context which in turn depends on the specific profile.

Various types of interfacing modules will be designed to support smooth communication of the intermediate server to the a/v archives. The major final product will be an integrated software system consisting of the two, *semantic unification* and *personalization subsystems*, together with two types of interfaces. Namely, those between the system and the individual a/v archives and those between the system and the end-users.

The project will produce novel tools and methods for extracting high-level semantic information through Dynamic Thematic Categorization (DTC) of a/v content units and Detection of Events and Composite Objects (DECO) within the archived a/v material. DTC will provide a fuzzy association of a/v units to the nodes of a Thematic Categorization Structure reflecting the interests of specific users/groups. DECO corresponds to the identification of composite objects or events that have semantic interpretation. A variety of associations will be possible at this level.

The project will reduce their number by exploring the specific user behavior and constructing "personalized" sub-associations; consequently, it will adapt its performance to the specific users' interests, thus increasing its information retrieval efficiency. Using statistics and relevance feedback will be examined to assist personalization. Hybrid intelligent techniques, mainly neurofuzzy approaches, will be explored for constructing/learning the most appropriate semantic associations.

3.1 Extraction of high level semantic information

One of the main research targets of FAETHON is the extraction of high level semantic information out of existing syntactic or (lower level) semantic data like those encapsulated in MPEG-7 structures (D's and DS's). Of particular importance to this function will be the subjective (depending on users' profile) extraction of semantic information (interpretation rules).

Current research and standardization activities in the area of a/v content handling and indexing are driven by the work related to MPEG-4 and/or MPEG-7. On the basis of (interacting) audiovisual objects, defined by MPEG-4, MPEG-7 introduces associated information structures of descriptive nature, i.e., provides a formal syntax for attaching information of various detail and nature to the raw a/v content. More interestingly this information is not only attached via textual characterization/classification of the a/v objects; it is also extracted -perhaps automatically- on the basis of visual and acoustic characteristics of the "signals". In addition, the spatio-temporal structure of a video document, i.e., the spatial/temporal evolution and the resulting relationships between objects are conveying rich semantics. Elementary Descriptors (D's) and the more complex structures of Description Schemes (DS's) are the containers of the aforementioned information. Although D's and DS's may carry both *syntactic* (color, names, dates, etc) and *semantic* (thematic classes, events) information it is clear that: (i) in terms of the MPEG-7 standardization actions, syntactic features have concentrated the major work force; (ii) the current state of techniques related to the definition and automatic extraction of semantic DS's is mostly related to the creation of *macroscopic indices* that point to temporal segments (scenes, shots, microsegments) of a/v sequences; (iii) MPEG-7 will probably standardize a few (more) general DS's of semantic nature but clearly will "assign" the definition of any other useful DS to the implementation of specific systems.

FAETHON project will contribute to the definition of semantic DS's mostly on the basis of already available D's and DS's. The evaluation of their values (the so called *description*) is intended to be *dynamic* in the sense that it will reflect the instantaneous *context* of queries, the latter being determined by the profile of the users. Hybrid tools of artificial and computational intelligence combined with neurofuzzy techniques will be employed to achieve this behavior. The DS's to be defined will be *flexible and operating on a (higher) abstract level* in the sense that they will be applicable to more than one a/v archives with possibly different internal descriptions and detail of description. Fuzzy rules will be used to overcome partial incompatibilities among the archives.

Through this very first objective FAETHON project intends to offer the capability of adding value to existing a/v archives by unifying their *view* in a *semantic level*. As described in the methodology section this may involve one or more third parties playing the role of service providers who will design, distribute and update the proposed semantic DS's.

3.2 User Profile generation and update

The second research target within FAETHON is the generation and update of user profiles in conjunction with the creation and handling of classification data associated to archived a/v objects/units.

The large data volume of a/v archives (and even more, of groups of archives, accessed via centralized brokers) may overwhelm users that attempt to search through their content. Recurrent refinement of users' queries may be necessary before reaching the desired a/v material. This results in the increase of the search time, the overload of the serving system and the possible waste of telecommunication bandwidth.

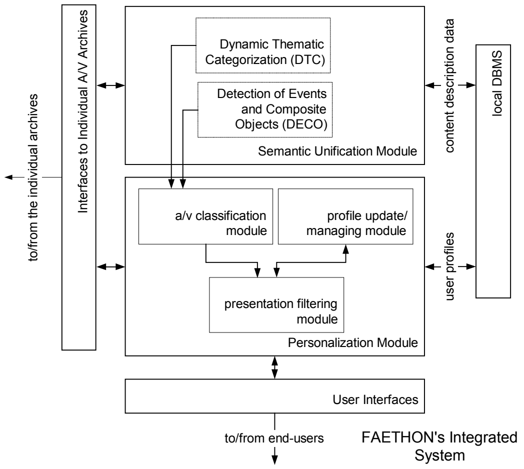
On the other hand, different users of the archived a/v content possess and/or set different priorities/preferences that could beneficially reduce the scope of their search or at least settle a ranking in the presentation of the responses.

The project will exploit this last observation implementing an intermediate server mechanism that keeps track of users' preferences, projects these preferences to appropriate indices of the archived content and adjusts the responses to users' queries in a manner that "fits" to their priorities. This subsystem will rely on three major components, namely an *a/v content classifier,* a *users' preferences tracking module* and a *presentation filtering module.* The a/v content classifier will retain a semantic information structure for the available content and its components. The previously referred semantic DS's will be dynamically evaluated components of the information structure. Users' preferences tracking module will monitor the choices of each user in order to trace his/her preferences. An important issue here is the representation and handling of the uncertainty that is conveyed in all the classification actions of user profiling and updating. The associated record, representing specific user's profile, will be continuously updated using neuro-fuzzy and relevance feedback techniques. Finally, the presentation filtering module will rank the responses of the archives to specific user requests according to the relevance of their "information structure" to the "profile" of the user.

3.3 Implementation Issues

The huge quantities of information produced by the tasks described in the previous two subsections, as well as, their increased complexity together with the need for storage and efficient retrieval of video documents and knowledge artifacts, lead to the design of an extended data base model that can handle this voluminous multitude of information. The data model will capitalize on existing data base architectures and facilities for extending them towards multiple dimensions and spatio-temporal feature handling.

In addition, since the proposed system offers "intermediate services", not replacing individual archives, its efficiency depends on the degree of achieved interoperability with the latter. FAETHON project will introduce a novel model of archiving architecture maximally cooperating with the supported services.


to/from the individual archives
Interfaces to Individual A/V Archives
Dynamic Thematic Categorization (DTC)
Detection of Events and Composite Objects (DECO)
Semantic Unification Module
content description data
local DBMS
a/v classification module
profile update/ managing module
presentation filtering module
Personalization Module
user profiles
User Interfaces
to/from end-users
FAETHON's Integrated System

References: