An MPEG-7 Compliant Integrated System for Video Archiving, Characterization and Retrieval

G. Votsis, A. Drosopoulos, G. Akrivas, V. Tzouvaras and Y. Xirouhakis

Image, Video and Multimedia Systems Lab.,
Dept. of Electrical and Computer Eng.,
National Technical University of Athens,
Email: `yvotsis@image.ntua.gr`

ABSTRACT

The archiving, characterization and retrieval of audiovisual material, stored in multimedia databases, are tasks of paramount importance mainly due to the large amount of information involved. Present work introduces an integrated system for video archiving, annotation and retrieval, developed for the Greek Ministry of Press and Mass Media. The particular system is made compliant with the MPEG-7 formalism and employs a graph-based approach to ensure data integrity.

1. INTRODUCTION

During the last decades, particular attention has been given to the manipulation and dissemination of audiovisual information, which has been for long unavailable outside the organizations where it is produced. Such organizations, e.g. audiovisual archives owners and TV broadcasters, are particularly interested in improving their techniques in both organizing their own audiovisual archives and in selling material via online services to external customers. In this context, distributed digital libraries providing advanced video annotation and retrieval mechanisms are highly appropriate. Systems incorporating indexing capabilities are also required to be open to accommodate standards refering to metadata like MPEG-7.

Given the emerging demand, significant effort is being made towards efficient storage and dissemination of audiovisual data. Research in this area exploits to a large extent the recent results in the fields of Database Systems, Computer Vision and Video Processing and Understanding. A number of prototype systems are nowadays in the stage of development as joint projects between academic forums and industries, or even in the stage of evaluation as commercial products. It is common that such systems basically ensure data security, storage efficiency and database integrity; often on the basis of some commercial database system. On top of the database system, advanced video annotation and video indexing mechanisms are developed to improve access and retrieval efficiency. The latter are rather dynamic modules of the system, which are constantly updated to comply with the significant advances in the field of Content-based Indexing and Retrieval. In fact, the extent to which such advanced capabilities are exploited, differs along with the scope and focus of the particular system. However, in general, efforts are made that prototype systems comply with the guidelines of MPEG-7, the latest Motion Pictures Expert Group coding standard.

Despite the great variety of video retrieval mechanisms and systems proposed in the literature, a few integrated systems world-wide agree with the aforementioned general framework, with AVIR [2, 12], VICAR [13], DIVAN [6, 11], ECHO [10] and VideoLogger [14] being some of the most common among them. VideoLogger [14] is probably up to now the most popular system as a commercial product, while ECHO [10] emphasizes also on the data gathering itself; the latter being a great portion of the historical european chronicles. In turn, AVIR [2, 12] concetrates on the effective coverage of the MPEG-7 standard.

In the present work, an MPEG-7 compliant video retrieval system, developed for the manipulation and dissemination of the historical archives of the Greek Ministry of Press and Mass Media, is described. The system, apart from its added-value in the fields of information characterization and retrieval, is of major cultural importance, as a pioneering work in the dissemination of historical data, gathered in film media from the beginning of the 20th century.

The proposed system employs a combination of manual and automated annotation schemes,

Figure 1: Snapshots

effectively interacting with each other. At the same time, it employs advanced content-based indexing mechanims for efficient access to the audiovisual information. In the particular system, digital information (such as video, images and text) is combined with non-digital data (such as films and tapes) in an MPEG-7 compliant hierarchy; adhering to the general notion that even a hand-written message can have an MPEG-7 description.

2. SYSTEM OVERVIEW

In the particular application, all information was basically available in film media kept by the Greek Ministry of Press and Mass Media since the beginning of the 20th century. In this sense, the only means of incorporating parts of the material in TV productions (e.g. documentaries) or expositions was through copying the original film or even re-writing it to a high-quality videotape (such as Digital BetaCam). However, this process has been proved inadequate due to the possible corruption of the original material, along with the fact that information can not be easily made accessible to the public. For these purposes, a system was implemented according to the following guidelines:

- provide support to different media types, such as different types of film, videotapes (e.g. VHS, BetaCam SP, Digital BetaCam) and digital video (e.g. uncompressed, MPEG-1, MPEG-2),
- ensure integrity and accuracy in correspondence of the same material in different media (even for full or partial copies),
- support copyright information for authenticated use of the material,
- facilitate partial reproduction of the material (e.g. portions belonging to a certain historical or political theme),
- provide means of effective manual annotation to experts on the field,
- enhance video partitioning in scenes and key-frame extraction prior to annotation,
- provide efficient mechanisms of browsing through database contents (both by literals and by visual example),
- comply with MPEG-7 guidelines for information management.

Within this context, the substructure that was required for the implementation of the system demanded work in three seemingly separate fields that should be harmoniously integrated with each other. A consistent digitization procedure would provide the support to different media types, while ensuring their accurate relevance. A semi-automatic annotation tool, in the hands of a knowledgeable group, would cover characterization needs under the prism of the MPEG-7 standard. Finally, a respectively functional database is designed in order to effectively manage all this information.

The digitization phase regards the conversion of the audio-visual material into a format, which is appropriate for efficient use via a digital database. One of the most common digital, compact format of audio-visual material nowadays is the MPEG, in any of its forms, i.e. MPEG-1, -2 and -4. For reasons, such as information compaction, speed of transfer via the Internet as well as compatibility with international standards, the developed system adopted these formats.

Concerning annotation, a tool has been implemented, which enables an expert to produce an MPEG-7 compliant hierarchy (regarding temporal and spatial segmentation), as well as to provide textual information, such as comments, chronology, event and person detection and/or identification. Wherever possible, the latter information is automatically extracted. However, the annotator has the authority to alter the results of the algorithms.

The system's physical architecture is depicted in Figure 2. In particular, there are five modules where information is stored (in fact, not all of these modules need to be physically separated).

In the *Film Storage*, *Tape Storage* and *Video Storage* modules, all audiovisual material is incorporated. In fact, these entities are not separated in terms of functionality, since they all

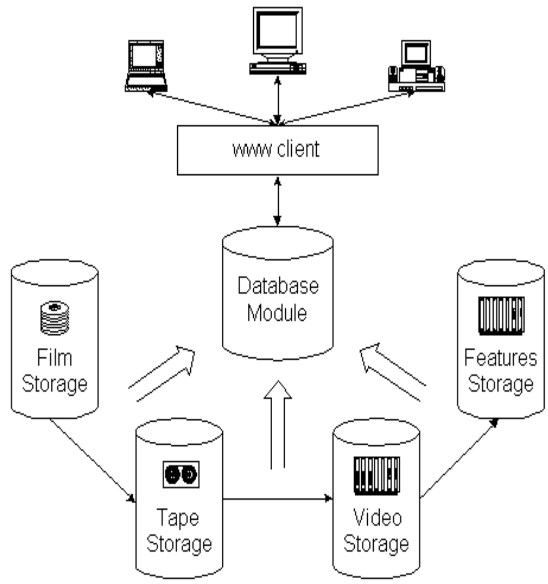

Figure 2: System Architecture

correspond to storage modules of a generalized medium type, however they correspond to different physical spaces; for example films and videos are kept into boxes and computer files respectively. At the same time, the flow arrows in Figure 2 illustrate the common paths of information flow: films are written to videotapes and in turn videotapes are digitized to produce video files. Since video files are often meant for distribution, an appropriate watermark is inserted after digitization.

In the *Feature Storage* module, all additional information extracted from the audiovisual material is included. In general, the module contains all data indexing information that provide efficient access to the material as, for example, partitioning information in film/video scenes, manual annotation of scenes, characteristic frames, content information extracted (feature vectors). In fact, the *Feature Storage* module can be incorporated in the general *Database* module, which is responsible to respond to all queries submitted by the *www client*. In turn, there is a wide variety of questions that should be aswered by the *Database* module; for instance retrieve all

- MPEG-1 videos containing the phrase "racing cars",
- *characteristic frames* containing human faces,
- *scenes* in any media belonging to *themes* related to war,

or even more technical questions, as

- propose the duration in timecodes (for BetaCam) of all *scenes* depicting primeministers of Greece from 1955 to 1965,
- retrieve exact position in video frames of all pictures resembling this one,
- depict the film length in meters and the film box number of a particular *scene*.

The above queries, submitted through the appropriate user-interface denote the underlying complexity of the *Database* structure. In short, the particular module is responsible for gathering all data, metadata, hierarchy and dependence information in such an effectively structured way, that even the most complex questions are answered fast. The tasks performed by the *Database* were in some cases particularly hard to fulfil when considering the following (common) examples. An entity recognized by an expert as a *scene* may reside in different parts of a film, or even in distinct films, or in general in different media. Any partial copy, or conversion from a medium to another must be accompanied by the semantic information already extracted for the first medium. On top of all the above, a new entity of data or metadata should at any time be possible to be inserted or deleted, without violating database integrity. In parallel, an open architecture was required so as to incorporate new features, as for example the expansion of video or image feature vectors.

In the following sections, appropriate details on the database structure and the respective updating mechanisms, the content-based indexing and retrieval schemes and the MPEG-7 framework of the particular system are pointed out and discussed.

3. EXPLOITED MPEG-7 ELEMENTS

The objective of the MPEG-7 standard (formally Multimedia Content Description Interface) is to provide a standardized way to describe the content associated with any kind of audiovisual information [15]. Content is represented by the features of the multimedia information, which are, in the context of MPEG-7, subsets of the data signifying something to somebody. The standard defines a hierarchical language, the Description Definition Language (DDL), based on XML, which can be used to represent the features using Descriptors (Ds). Descriptors resemble the records of a database, and can be composed of simple data types, such as numerical and alphanumerical. Simple data types are explicitly defined by the standard. Descriptors are hierarchically organized by Description Schemes (DSs), which can contain other DSs and/or Ds. The language enables a system to construct a hierarchy of DSs, which can describe, uniformly, many kinds of audiovisual information, both digital and analog.

Motivation for creating the new standard came from the realization that, due to the digital revolution, audiovisual content is increasingly consumed and processed by machines, for example in multimedia databases (efficient retrieval of multimedia content of interest to the user) and various kinds of medium understanding, e.g. computer and robot vision. Machines need a higher level of structure than humans, in order to perform these kinds of advanced operation. Attempting to fulfil these requirements, the standard aims to deal with the problem of interpreting the multimedia content, in a way that can be automatically processed by the machine.

Just as MPEG-4 provides a standardized way to describe the objects composing a scene, without providing the algorithms needed to extract the objects, the MPEG-7 standard does not provide the algorithms needed to extract the description. For some of the features, automatic extraction can currently be achieved, for example scene cut detection [16], or it will probably be achieved in the near future, such as object identification, but for some of them, automatic extraction without human intervention is very hard, or even impossible; for example extracting semantic and context information, annotation of the various *scenes*, giving chronology of the events, etc.

Because the MPEG-7 standard is not yet ready, one cannot ensure that the currently published details of the standard will be retained. The attractiveness, however, of many of the published features of the standard, specifically the hierarchical representation of the content of the audio-visual information, urges for them to be exploited into the system described in the present work. We now describe the MPEG-7 features used in the system.

The highest level of the DS hierarchy is shown in Figure 3 in UML format. Audio visual (AV) content is represented by the following DSs:

- SyntacticDS: segmenting an AV entity in space (e.g. object extraction) and time (e.g. scene detection)

- SemanticDS: specifying semantic features, such as objects, events and the relations between them

- MetaInfoDS: specifying information, mostly literal, about the medium, such as annotation and author information

- MediaInfoDS: linking AV content to the physical medium, either digital or analog, where it is stored

- SummarizationDS: using a selection of frames and scenes to enable fast browsing of the content

Apart from the above DSs, the standard defines a ModelDS, which is used to support synthetic models. However, since our system only deals with natural video, we decided against supporting synthetic models.

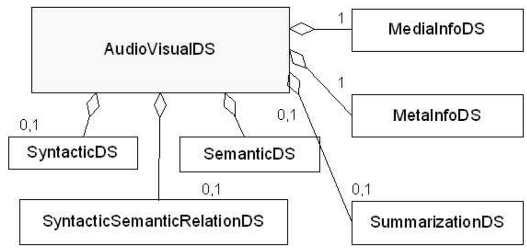

Figure 3: MPEG-7 DS

SyntacticDS contains all segmentation information, both in time and in space (Figure 4). This is achieved through a generic DS, Segment DS, which refers to any segment of AV information. A segment is associated, through a MediaInfoDS, with the media where it is stored. Movind and Still RegionDS represent kinds of spatial segmentation, while VideoSegmentDS supports temporal segmentation. Our system supports 3 levels of temporal segmentation, namely *scene*, *scene group* and *theme*. In the MPEG-7 implementation, the annotator is able to exploit and expand the output of a scene cut detection tool, while being able to organize the *scenes* in the 3 levels. By using a graphical user interface, regions of interest are defined by the annotator. Also, the segments are linked to semantic entities, such as objects and events, by the annotator.

MediaInfoDS links a segment of the AV information to its physical storage (Figure 5). This is very important to our system, because of the multiple media involved. For each medium (for example, film, videotape, digital video), MediaInfoDS contains a MediaProfileDS, which contains all the necessary information for the medium, the segments location in the medium. This part of the system must employ a high degree of automation. Particularly, it must be able to map

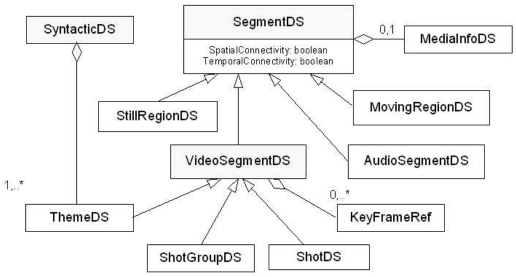

Figure 4: SyntacticDS

the time boundaries of the various segments from one medium to the others. Because of the particularities of the material used in our system (segmentation of scenes into several locations on the film), the standard MPEG-7 scheme needs to be expanded, to enable medium segments to be defined by the annotator.

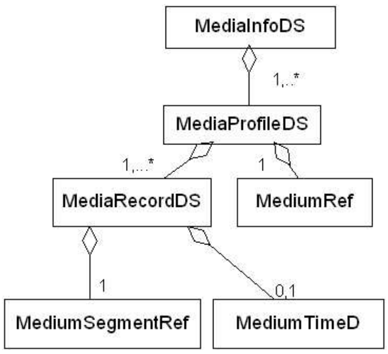

Figure 5: MediaInfoDS

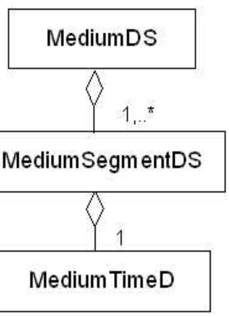

Figure 6: MediumDS

The MetaInfoDS deals with the literal information (metadata) created by the annotator and associated with the audiovisual data. This is of great importance for our system, due to the historical importance of the archives. Our system incorporates the classification in categories that was already used by the archive owner, but also supports its expansion by the annotator. An example of this scheme is shown below.

SummarizationDS is going to expand the keyframes defined by the annotator. According to the MPEG-7 standard, the characteristic frames are organized hierarchically. Moreover, whole *scenes* or higher level segments can be used to summarize the signal.

4. SCENARIO AND IMPLEMENTATION DETAILS

4.1. Digitization procedure

As mentioned in section 2, all available audiovisual material had to be digitized in order to be efficiently used by the database. In our case, the original material was in negative 16mm film format, including parts dating even 90 years. Its low material quality required special pre-processing in order for the content to be finally converted to a video signal, through the employment of a digital Telecine, into a digital Betacam format.

To aid the temporal segmentation of a video signal into semantic time segments such as *themes* and *scenes*, MPEG-1 and MPEG-2 files were initially created for each film roll, through the procedure described in Figure 7. The video signal was encoded via the Matrox Digisuite digitization system, by the Optibase MPEG Fusion encoder with frame accuracy. The optical signal had PAL resolution ($720 \times 576 \times 25$), while the sound bitrate was adjusted to be compliant with MPEGs (48000×16). The MPEG signal was generated with a target bitrate of 5 Mbps as far as the MPEG-2 Full D1 resolution is concerned, while in the case of MPEG-1 SIF resolution (352×288) a target bitrate of 1.6 Mbps was selected. During the digitisation procedure, the Matrox Digisuite overlaid the required logo at the bottom right corner of each frame.

These initial MPEG files were handed over to the annotator. The latter actually performed the segmentation procedure and of course the annotation by means of the respective tool, which is described in subsection 4.2. On the basis of the temporal data that arise by the characterization procedure, the video signal was re-digitized to meet the needs of the time segmentation, i.e. each segment forming a *theme* defined a new, independent MPEG file. This procedure also produces detailed information about the relation between digital Betacam timecode and MPEG timecode.

Due to technical or historical reasons, a segment defined as *theme* could often be scattered in diverse pieces of the original films. That is mainly the reason why the procedure described in Figure 7 could not be followed in order to digitize the new, briefer pieces of visual information. The process, which was in fact followed, is depicted in Figure 8 and elaborated in the following.

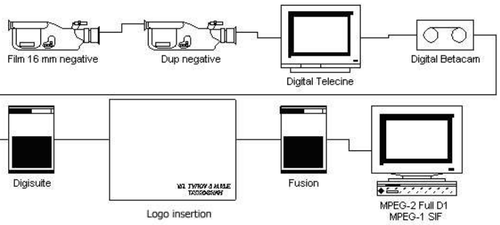

Figure 7: Film-roll to videotape conversion

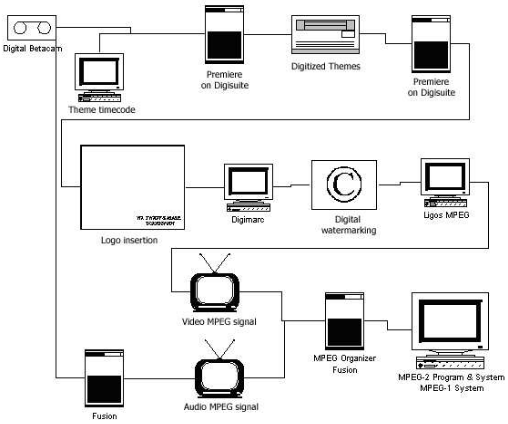

Figure 8: Videotape to MPEG-file conversion

The preliminary digitization and concatenation for the formation of each *theme* was executed by applying Adobe Premiere, which had been fine-tuned previously by Matrox in order for the former to cooperate with Digisuite. The output of the first step was in AVI format, while Premiere was used to overlay the logo in the bottom right corner for the whole of each *theme*. Subsequent to this phase, watermarking was additionally applied to the sequence with the use of Digimarc Embedder. By running the application Ligos MPEG Encoder, MPEG encoding followed the watermarking procedure. Three different types of MPEG files were produced for each *theme* from this subsystem: PAL Full D1 (720×576), SIF (352×288) and QSIF (172×144).

Encoding proved difficult due to the fact that the original sequence was cinematographic. Such signals, especially those that date from the dawn of the 20th century, are characterized by great fluctuation in the image value (if one considers the signal in HSV representation), by abrupt frame displacements, as well as by various other noise sources. The MPEG encoder is forced to introduce relatively large errors in order to maintain the bitrate within limits. To restrain this phenomenon, bitrates were selected to be somewhat higher than usual.

Due to the fact that sound was not watermarked, it was directly digitised though Fusion. The final audio and visual signals were multiplexed with the use of Fusion in order to produce System and Program MPEG sequences.

Digital watermarking was judged to be a crucial part of the procedure as a sophisticated way to assure ownership rights. To develop watermarking methods that are resilient to various deformations, which the images of a video sequence might undergo (such as redigitization, noise addition, image cropping), is a matter of research. Furthermore, techniques for directly embedding watermarks in MPEG streams are being examined, but they are also still in the stage of research [17].

The watermarks intensity is one of the tradeoffs between video quality and watermark endurance. In our case, as described in section 2, the video signal had to be MPEG encoded, a fact that forced towards high watermark intensity. The selected, high value finally bestowed a mean SNR of 25,8 dB, while the watermark endurance percentage varies from 10% (1 in 10 frames remained watermarked after decoding) - in the case of old, low quality movies- up to 90% in the best case. These percentages retain fairly enough the objective of the watermarking procedure. However, low percentages are mainly due to the frame-by-frame practice of the method. On the other hand, direct application of a watermarking mechanism upon an MPEG sequence should provide a solution to the problem. Let us point out that existing commercial products remain feeble in this sense.

4.2. Syntactic analysis and MetaInfo generation

The annotation application provides its user with the means to browse a part of the archive, in MPEG format, in order to determine and segment logical entities in it and then add textual information to them. To minimize the effort and time needed for manually extracting the related information (moments of scene change) was the main motive for implementing an automatic scene change detection mechanism. *Scenes* are defined, in the present system, to contain the audiovisual content captured by a camera from the moment the later starts to shoot until the moment it stops or changes its viewpoint to capture a new *scene.*

The system must, ideally, find the moments of various kinds of scene changes (abrupt changes, gradual changes, fade ins, fade outs etc.) rapidly, while keeping erroneous detection (non-existent changes and detection failure) as low as possible. The system implemented consists of two parts, one for scene detection, and one for enabling the annotator to examine the results, and possibly correct them.

For scene detection, our mechanism uses an algorithm proposed by Yeo and Liu in [16]. This algorithm has the advantage of using the MPEG signal directly, making full MPEG decoding unnecessary. Processing is quite fast, approximately real time for MPEG-1 resolution on a Pentium III PC. Error is around 3% for false scene detection and around 2% for failed scene detection for color TV signal. However, error for the videos involved in our system was somewhat higher, because not only were the videos Black and White (giving only one color channel to process instead of 3 for color videos), but also old cinema movies, with all the problems mentioned in the digitization section.

The output is a series of frame numbers, corresponding to the scene changes. The output can be viewed by the annotator, who can then alter the scenes found by the system. The data is finally stored into the system (syntactic part, Videosegment analysis.).

Apart from detecting scenes, the algorithm proposed in [16] can also extract characteristic frames. As mentioned in section 3, a set of characteristic key-frames constitutes a sequential summarization. This is stored in the summurization part of the system. Moreover, with scenes and key frames detected, the relevant annotation interface enables the annotator to built a hierarchy of selected scenes and key frames.

Hierarchy structures are useful for providing the user a fast way to examine a video (video browsing), and also for video query and categorization applications [1].

Once *scenes* are detected, the annotator views them and organizes them into *scene groups* and *themes*. A *theme* represents a logical entity in the medium. At this point it should be noticed that there is no constraint for the *scenes* of a *theme* to be part of the content included in a single physical medium (Figure 9).

In order to provide a more detailed hierarchy of the *themes*, the system includes the de-

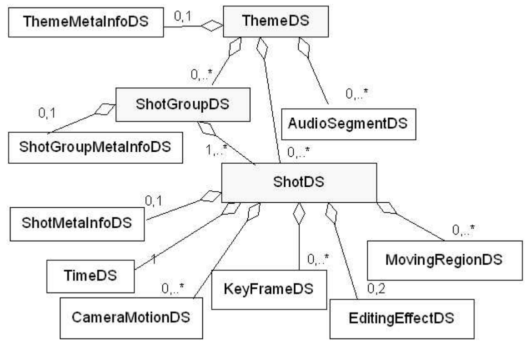

Figure 9: ShotDS

finition of *categories* and *subjects* under which *themes* are organised. *Categories* are predefined groups a *theme* can become part of, based on its logical content. *Subjects* are formed in the same way, differing only in the fact that the annotator defines them at the point of his work where he finds need for them, and that one *theme* can be part of one *subject* group and one or more *category* groups.

Once the annotator has determined the physical boundaries and the hierarchy of the content based on its semantic information, it is then his task to provide the system with textual information describing those semantics. This is performed in various points of the annotation procedure. At first the annotator characterizes the *theme* entity providing basic information such as the date, historical period, and place of the shooting along with a general description of the content. If possible to be determined, information is added concerning the people that took part in the shooting or in any post-shooting tasks for the current *theme.* All this information provides a powerful categorization and description of the archive content. Along with this, the annotator can characterize a number of *characteristic frames* of a *theme* by supplying additional textual description becoming another form of *theme* representation in the archive. Finally, annotation is also performed for the different *scenes* that form a *theme* in order to highlight significant historical information that can be included even in a small number of frames in such an archive. In addition the annotator can insert textual information for *groups of scenes.*

4.3. Automatic feature extraction

Segmentation is not only performed in the temporal dimension of the information by the annotation tool, but also in its spatial dimention. So, in the proposed system, a number of content-based indexing schemes have been included. As in most systems (see for example Virage [5], QBIC [4], Photobook [19], VisualSEEk [20] among many others), an open architecture has been followed in order that such functionalities are easily extended w.r.t. the recent progress in this field. It must be pointed out that not any image attribute can help to the direction of image retrieval, since extraction and similarity measures efficiency generally depends on the data itself; for example in a database containing underwater scenes a global color criterion would probably fail. Regarding the particular database, the following attributes were proved adequate

- *face detection:* detecting possible presence of human faces in video clips, since most people included in this type of audiovisual material are expected to have an important political and/or cultural role,
- *mobile object detection:* main mobile objects are expected to include semantic information that can mainly assist the annotator in the localization of frames of interest in the clips,
- *composition:* since a large percentage of videos do not contain color; an intensity composition measure is preferable to a global color measure for comparing characteristic frames (key-frames).

Figure 10: Face detection results

Face detection: Various algorithms have been developed to deal with the face detection task. Such algorithms are the template-based detection, feature-based detection [18] and neural-based detection. The proposed algorithm uses features from the latter algorithms. In addition, an additional strategic is proposed, which accelerates the process. This strategic adds on the speed of the algorithm, which -for such a real application- is considered to be one of the most important factors. According to this, one may discern two stages in the proposed algorithm, the slow one and the fast one. It is known that the ratio of the vertical to the horizontal dimension of a face is $\frac{5}{4}$. In a picture with a dimension of 576×720 we have $(576 \times 5N) \times (720 \times 4N)$ possible faces. If N takes value from $4 - 12$ then we have a total of 74500 faces in the picture. In its first stage, the algorithm rejects 70 percent of the possible faces using the information of the dimensions ratio of the face in addition with the amount of luminance of the face. A database of images was used to statistically calculate the variables refered to the luminance of the face or other information that was needed for the development of the algorithm. This database consisted of 40 faces in 10 different poses each and in various luminance conditions. These images were selected to cover all the possible combinations of various extra characteristics, such as glasses, hair, moustache. The second stage of the algorithm examines the remaining candidate faces in greater detail. The face is split in three areas; the area of the eyes, the nose and the mouth. An edge detection algorithm is applied in each of the three areas. The experimental results are shown in the following Figure 10.

Mobile object detection: Mobile object detection schemes involve the detection of main (large) mobile objects in the scene and their localization on the image plane. Such schemes either focus on extracting the mobile object contour for object classification and/or recognition, or treat the object by its coarse bounding polygon. The former have been proved to provide poor classification results in the general case, although they involve significant computational cost. On the contrary, The latter, being particularly faster, aim to extract descriptions like 'a mobile object covering 20 percent of the image plane moves to the right image side (specifying a certain angle of direction)'. The description can well be more detailed according to the features extracted for the particular object, such as area, centrality or curvature. In our case, following the MPEG-7 approach, the bounding polygon proves sufficient for both the annotator and the user, so that mobile objects are easily retrieved and handled as separate regions of interest. For this purpose, we employ the algorithm presented in [8] relying on transient image edges to extract mobile regions. The particular algorithm is highly parametrized to form succesfully the objects' bounding polygons for both indoor and outdoor scenes for static or mobile camera capture.

Figure 11: Mobile object detection results

Color composition: The color composition image attribute has been widely treated as the most popular criterion for the simultaneous extraction of color and shape information (often termed as 'composition'). In this case, the graph-based approach of [9] was implemented in order to retrieve characteristic frames similar to an example one, in the sense of 'querying by visual example'. The user can then efficiently browse through different clips depending only on the visual similarity of their respective keyframes. The employed algorithm is functional even when only certain parts of the images under comparison are similar, or when images are obtained under different affine transformations (as this can be due to camera motion even in the same clip).

4.4. Media Linking

The database module forms the backbone of the presented system, as it determines its structure and the relationship between its features while it is, at the same time, responsible for the circulation of the semantic content information between the different storage media. The design of the database scheme determines the hierarchy followed in the audiovisual content representation and, as a result, its MPEG-7 compliance. In addition, the database architecture should follow an adequately abstract and flexible model in order to be addressed for various tasks, that could emerge, such as producing reports of the current archive status, and to be open for expansions of the system that might occur in the future.

As described in 4.1 the audiovisual content consists of an archive stored in film which is duplicated in video-tape and finally reproduced in MPEG format. That means that the database holds information for the different media types in which the same content is stored, possibly including safety or commercial copies that are produced. Using an appropriate application the archive administrator updates the database to contain all the information derived through the digitization procedure in order to ensure correspondence of annotation information. In this way, the system contains sufficient information to form a graph whose nodes are the different physical media where archive content is stored (Figures 5,6). The graph edges represent the relationship between parts or the whole of these media providing in this sense essential information about the origins of the contents of any physical medium produced in the digitization procedure. It should also be noted that the database contains enough information to determine the direction of the graph edges in order to provide information about the origin of a medium's content.

As shown 4.2, the use of the annotation application provides the database with a significant amount of physical and semantic information that characterises the archive's content. This information is produced based on a digital video which the annotator browses and semantically determines. It is though already noticed that this content can exist in various other physical media that are also basic parts of the archive. In this context it is essential for the system to be able to pass the hierarchy formed and the additional information provided about the content to all the other storage media too. This is achieved based on the graph representation of the medium information that the database contains. Thus, the results of one of the annotator's tasks which provide the database with new *scenes* and *themes* in a digital video, along with textual information about them, circulate in the virtual graph in order to be associated with the same content in any other medium type.

This can become a quite complicated task since in a large archive there are many duplicates of the original media and even more storage media that contain only some parts of one or more others. Here the graph approach used for media linking proves its efficiency since it can be determined in a straight-forward way whether a medium should receive information about its content by another one or not. As it can be noticed this approach also supports the update of the whole database when editing or even deleting semantic information about the content in a physical medium retaining the integrity of the system.

5. CONLUSION

In this work, an integrated system for video manipulation and retrieval is presented. The proposed system was developed to support the historical archives of the Greek Ministry of Press and Mass Media. Particular attention is given so that metadata information is compliant to the MPEG-7 standard.

Acknowledgements: The present work is supported by the project "Digital Archiving of Audiovisual Data of the Ministry of Press and Mass Media", funded by the Greek Ministry of Press and Mass Media, 1999-2000.

6. REFERENCES

[1] Y.S. Avrithis, A.D. Doulamis, N.D. Doulamis and S.D. Kollias, "A Stochastic Framework for Optimal Key Frame Extraction from MPEG Video Databases," *Computer Vision and Image Understanding* **75** (1/2) (1999) 3-24.

[2] M.Ceccarelli, G. Mekenkamp and E. Persoon, "AVIR - Home Multimedia Systems: On Personal Video Libraries," *ICMCS'99*, Florence, Italy, June 1999.

[3] S.-F. Chang, W. Chen, H.J. Meng, H. Sundaram and D. Zhong, "A Fully Automated Content-Based Video Search Engine Supporting Spatiotemporal Queries," *IEEE Trans. on Circuits and Systems for Video Technology* **8** (5) (1998) 602-615.

[4] M. Flickner *et al*, "Query by Image and Video Content: The QBIC System," *Computer* **28** (1995) 23-32.

[5] A. Hampapur *et al*, "Virage Video Engine," *SPIE Storage and Retrieval for Image and Video Databases V*, vol. 3022, pp. 188-198, San Jose, CA, 1997.

[6] A. Tirakis, P. Katalagarianos, M. Papathomas and C. Hamilakis, "Distributed Audio-Visual Archives Network," *ICMCS'99*, Florence, Italy, June 1999.

[7] G. Votsis, Y. Xirouhakis, K. Karpouzis and S. Kollias, "Efficient Use of Content-Based Retrieval in Quick Browsing: A Realization," *MMSP 1999*, Copenhagen, Denmark, September 1999.

[8] Y. Xirouhakis, V. Mathioudakis and A. Delopoulos, "An Efficient Algorithm for Mobile Object Localization in Video Sequences," *VMV 1999*, Erlangen, Germany, November 1999.

[9] Y. Xirouhakis, A. Tirakis and A. Delopoulos, "An Efficient Graph Representation for Image Retrieval based on Color Composition," *CSCC 1999*, Athens, Greece, July 1999.

[10] ECHO, "European CHronicles On-line", http://www.ina.fr/Recherche/Echo.en.html

[11] DIVAN, "DIstributed audioVisual Archives Network", http://divan.intranet.gr

[12] AVIR, "Audio Visual Indexing and Retrieval", http://www.extra.research.philips .com/euprojects/avir/

[13] VICAR, "Video Indexing, Classification, Annotation and Retrieval", http://iis. joanneum.ac.at/vicar/

[14] Virage VideoLogger, http://www.virage. com/products/videologger.html

[15] "MPEG-7: Context, Objectives, and Technical Roadmap," ISO/IEC JTC1/SC29/WG11/ N2861.

[16] Boon-Lock Yeo and Bede Liu, "Rapid Scene Analysis on Compressed Video," *CSVT 1995*, Vol. 5, NO. 6, December 1995.

[17] F. Hartung and B. Girod, "Copyright Protection in Video Delivery Networks by Watermarking of Precompressed Video," *Springer Lecture Notes in Computer Sc.*, Vol. 1242, pp. 423 - 436, 1997.

[18] K. Karpouzis, G. Votsis, N. Tsapatsoulis and S. Kollias, "Compact 3D Model Generation based on 2D Views of Human Faces: Application to Face Recognition," *Machine Graphics and Vision*, vol. 7, nos. 1/2, pp. 75-85, 1998.

[19] A. Pentland, R.W. Picard, S. Sclaroff, Photobook: Content-Based Manipulation of Image Databases, *International Journal of Computer Vision* **18** (3) (1996) 233-254.

[20] J.R. Smith, S.-F. Chang, VisualSEEk: a fully automated content-based image query system, *ACM Multimedia '96*, November, 1996.