A Multimedia Database Workbench for Content and Context Retrieval

Cristina Ribeiro
FEUP/INESC-Porto
Rua Dr. Roberto Frias s/n
4200-465 Porto, Portugal
Email: mcr@fe.up.pt

Gabriel David
FEUP/INESC-Porto
Rua Dr. Roberto Frias s/n
4200-465 Porto, Portugal
Email: gtd@fe.up.pt

Catalin Calistru
INESC-Porto
Rua Dr. Roberto Frias s/n
4200-465 Porto, Portugal
Email: cmc@inescporto.pt

Abstract—Multimedia databases are extending the scope of traditional databases to handle the complex structure of multimedia objects. Models for multimedia information must include representations for the structure and content of several media in a form that allows flexibility in retrieval. Content-based retrieval is the main motivation behind recent research in multimedia databases. The task of searching in video and audio content is made hard by the nature of audiovisual data where, unlike text, there is no direct syntactic channel between the object and its meaning [1]–[3].

We propose a model for multimedia content storage and retrieval accounting for both context and content information and taking advantage of their dependencies for effective retrieval. We then describe a prototype multimedia database with a retrieval interface. It has been used as a workbench for testing the representation model and integrating tools for feature extraction, information interchange and retrieval. The workbench allows an easy integration of new tools for content analysis and new methods for context- and content-based retrieval while offering storage and access for both the actual digital content and its metadata.

Keywords: Multimedia databases, multimedia content description, video analysis, content-based video retrieval, semantic gap.

I. CONTEXT AND CONTENT RETRIEVAL

We argue that a model for multimedia content storage and retrieval must account for both context and content information, taking advantage of their dependencies for effective retrieval. It is also important to be aware of the main concepts embodied in the model for capturing the nature and features of the represented objects. This can be correctly accounted for by adhering to the standards that are accepted by the audiovisual content producers and by the communities concerned with using, archiving and preserving such content.

A. Units of Description, Segments and Descriptors

Two main concepts are associated with the objects to be described. The first is the Unit of Description (DU), capturing the notion of an object or collection of objects that can be given a context in terms of creation and relationship with other objects. The second is the Segment, capturing the notion of some part of an object that can be independently analysed in terms of content and be used by itself. A segment has no context of its own, getting it from the Unit of Description of the object it belongs to.

In terms of structure both DU’s and segments are organised as part-of hierarchies. A DU is either a root unit or is related to the DU describing the collection of units where it belongs, which has its own DU. The same applies to a segment, which may be a part-of another segment. Any item that can be individually retrieved has an associated context, either the one in its own DU or that of the object including it.

Another basic concept for the description of content is that of a Descriptor. The sense in which Descriptor is used is the one established by the MPEG-7 Standard—a representation of a feature.

A Descriptor Value is an instantiation thereof. The set of descriptors is open in the adopted data model, so that new descriptors can be incorporated without modifying it.

B. A Data Model for Multimedia Databases

In the data model, the DescriptionUnit class captures the concept of a DU. An instance of DescriptionUnit is linked to an instance of a higher-level DescriptionUnit, in a many-to-one relationship.

Metadata concerning object content is centred in the Segment and Descriptor classes. A segment may be linked to some DU providing its context. Several segments may be attached to the same DU, accounting for different parts of its content. A segment may instead be attached to another segment, providing a specialised description of one of its parts. Segments are further specialised for text, image and video and are linked to descriptors specific to each media.

Fig. 1 shows the classes and relationships for DescriptionUnit, Segment and Descriptor. The model is able to incorporate the standard MPEG-7 descriptors as well as any user-defined descriptors.

The metadata model has been translated to an operational database system where tables and attributes closely follow the proposed model. This relational prototype has been populated with 4 sets of data: descriptions of documents from an historic archive, a photo collection (both descriptions and digitised versions) and video and audio from the MPEG-7 test sets.

II. THE VIDEO ANALYSIS FRAMEWORK (VAF)

The Video Analysis Framework (VAF) is an application that provides an integrated environment and a convenient interface for experimenting with video analysis tools. The VAF offers an object-based visual interface for configuring and running analysis tools. The central concept is that of an “engine” that performs a specific task or several related tasks such as data storage, data flow or data processing and is graphically represented by a rectangle with an associated name. Control and data information are exchanged between objects using messages, graphically represented as line connectors. The VAF supports operations such as video parsing and video representation.
with different levels of abstraction (scene, key-frame and objects), each one with its own video engine. A graphical user interface (GUI) allows users to select tools, assemble their processing graph and determine the flow of information.

Fig. 2 shows some aspects of the VAF interface. The main window contains a video graph where the “file 0” block feeds three parallel video engines. From the top, the first engine outputs its own input, the second is an edge detection engine and the third is an engine for extracting the MPEG-7 DominantColor feature. The outputs are connected to testpoints whose states can be observed in the results window, on the right. Its upper left corner presents one frame of the original sequence, the lower left part corresponds to the edge detection results and the right side is the output of the MPEG-7 DominantColor extraction.

III. ARCHITECTURE OF THE WORKBENCH

The architecture of the proposed workbench comprises three major components: the VAF, the database and the Search/Upload interface.

A number of video engines which output MPEG-7 compliant descriptions are already integrated in the VAF. These include DominantColor, MotionActivity, FaceDetection and AudioClassifier. Others like SpeakerIdentification, FaceRecognition, Scene Cut Detection, are working in VAF but are not MPEG-7 compliant yet. The descriptions produced by these engines along with the corresponding media are stored into the database through the VAF’s database module [4].

The second component of the workbench is an Oracle Database Server with XML capabilities and conforms to the data model described in Section I-B. The database can be populated either directly by the VAF or by the Search/Upload Interface. We make use of the BLOB (Binary Large Object) type for storing the media content and of XMLTYPE for XML descriptions.

The third component, the Search/Upload Interface, is used for searching and retrieving information from the database and for uploading both multimedia objects and associated descriptions. It uses the Apache struts architecture and has a three layer design separating the user interface from the controller that validates the data and identifies the actions to be taken and from the business logic layer which interacts with the database component.

The upper left part of Fig. 3 shows the VAF GUI (1). The lower part depicts the database component (2) with the hierarchy of description units and segments. The upper right part presents the Search/Upload interface (3) to be detailed in the next section. (1) communicates with (2) for the upload of content and generated descriptions. (2) and (3) communicate for content upload and retrieval.

IV. MULTIMEDIA RETRIEVAL

Search in the contextual metadata has two modes. In the direct mode, each DU record satisfying the criteria using its own field values is retrieved. In the inheritance mode, also the DUs that satisfy the criteria using field values inherited from upper levels (remember that the DUs constitute a part-of hierarchy) are retrieved. Typically, the result set is organised as a set of trees of DUs. In both modes, it is possible to traverse the result set and, at any point, to enter into a browse mode where one can navigate to sibling, parent or descendant DUs. This mode is specially useful for occasional users who do not know exactly what is in the database.

If several criteria are simultaneously active, the interface allows the establishment of the intended relative importance by the indication of weights for each one, in percentage. This linear combination of distances is used as a global distance.

V. CONCLUSIONS AND ONGOING WORK

The workbench is based on a multimedia database with a hierarchic structure and a clear integration between context and content metadata. Tools for feature extraction, object upload and information retrieval communicate through the database and remain open to experimentation.

Current work includes the validation of the proposed metadata model in a large database, the development of new criteria for searching on multimedia content and the use of the information gathered in the database to export complete MPEG-7 descriptions including information on context attributes along with the existing MPEG-7 descriptors.

REFERENCES