A Semantic Annotation Tool for Educational Video Retrieval

Maria de Fátima Webber do Prado Lima¹, Joel Luis Carbonera², João Luis Tavares da Silva³

¹ University of Caxias do Sul, Brazil. mfwpadima@ucs.br
² University of Caxias do Sul, Brazil. joelcarbonera@gmail.com
³ University of Caxias do Sul, Brazil. jlttsilva@ucs.br

Abstract: In this paper, we propose a framework for a multimedia retrieval system based on the MPEG-7 Metadata Schema. The framework goal is the high-level semantic retrieval, which is composed by three modules. This framework provides a rich set of manual feature components and an independent retrieval interface.

Keywords: Collaborative Learning, Multimedia, Networks, User Groups,

1. Introduction

With the increasing variety of multimedia content available and the increase computer storage capacity, the use of video data is becoming a teaching resource that can contribute to teaching qualification. This new form of content presentation tends to be widely used in educational environments. It facilitates the content assimilation and handling. The multimedia content is handled as a book. The user can search for a particular content, go forward, stop and go back in a simpler and quicker way. These features make the presentation much more attractive to the student and much easier to be handled by the teacher.

Unlike the textual content that is easy to index, the multimedia content cannot be searched in the same way, and to create methods to facilitate the search. Therefore, it is necessary for educational environments incorporate tools that can handle multimedia content. These tools should incorporate techniques to handle video data storage, transmission, indexing, and retrieval. The multimedia retrieval applications are much more difficult to model and to extract content than the text-based documents retrieval systems, because of the richness of audiovisual information content.

The video content indexing can be performed using two types of features: low and high. Low-level features [1] include color, shape, texture, motion, spatial
information, symbolic strings (visual features), waveform and power, spectral content, and timbre (audio features). High-level or semantic features are the ones related to the perceived meaning, video subject, context objects or scenes among others. These semantic features can be handled by different detection mechanisms that segment the multimedia material into regions of interest and attribute semantic annotation to each region and their relationships.

This work constitutes an indexing video tool that uses high-level features to select video segments through of some keywords that give meaning to the selected video. Furthermore, it provides a search engine that recovers video excerpts through keywords. This tool is being developed as a component in an educational portal, and it is being designed in a client/server paradigm. The teacher may include videos, select snippets of video and create annotations on the server machine. Students, using the client view, can watch videos and search for video snippets through keywords. For example, teachers in the health area could include videos showing different surgery stages. Using the tool for indexing, they could index the various stages of surgery, or link scenes to show the correct use of the surgical instruments. Students use these pieces of video previously marked by the teacher, facilitating the understanding of medical procedures.

Our video annotation tool is part of a larger effort to apply web technology, multimedia content and virtual learning environments in conjunction with the Health Department at UCS in a project that aims to update the teaching of health and Medicine courses.

The standard MPEG-7 [2][3] was used to develop this tool. MPEG7 describes the features of multimedia content so that users can search and retrieve content in a way similar to what occurs in the search engines on the Internet. In order to create these descriptions, the standard MPEG-7 offers a set of metadata elements described in XML (eXtensible Markup Language). This standard provides a good set of tools for standardization and description of multimedia content and also allows for inclusion of events, pictures and colors annotations of the elements and also their structures. These annotations are called descriptors that allow a quick access to the desired information.

The organization of the paper is as follows: Section 2 discusses some related work on video semantic indexing systems. The Section 3 presents some existent tools for video annotation. The semantic video content model is presented in Section 4 and also our video annotation tool for educational environments. Finally, Section 5 concludes the paper.

2. Related Works

There are many works on video indexing. Matos [4], Shao [5], Allassia [6] and Bailer [7] are concerned with low-level features. Matos [4] proposes an
application for automatic summarization for generic audiovisual content based on a MPEG-7 compliant hierarchical summary description. Matos [4] selected three low-level features: motion intensity, the density of shot cuts and the sound energy. Shao [5] has proposed an image retrieval method based on MPEG-7 which extracted eight dominant colors of each image and generates a histogram intersection to measure similarity. Allasia [6] describes a framework to extract low-level features using image and speech analyzers. Bailer [7] has defined a general description scheme for visual impairments (but not the extraction method). The suggested descriptors for visual impairments are defined under the next information: dropout/partial frame damage, line scratches, number of line scratches, dust/dirt level, noise/grain level, flicker level, image instability, blur, blocking level, dropout level, channel misalignment and color range defects.

Several video indexing tools have been proposed using the MPEG-7 standards. Kim [8] wrote that MPEG-7 standards is still missing licensing conditions because of the industry have not fully incorporated this powerful and yet unexplored standard. Nevertheless, many works have studied several important aspects on the development of tools for video indexing. Kim [8] proposed the customization of the multimedia content description MPEG-7 standard and the MPEG-21 multimedia framework. Gruhne [9] describes a cross-modal search based on the MPEG query format, using two different types of databases and a provider service to distribute the queries and aggregate the results. San Miguel [10] describes a distributed framework for real-time video analysis. The framework generates a MPEG-7 description that is stored in a database. The distribution of the video is based on a server/client model where the client logs on the server and data transfer is started. The flow control is realized through a TCP-based network.

Ma [11] proposed a framework for multimedia retrieval system based on the MPEG-7 standard. The framework provides a set of automatic feature extraction components and an independent retrieval. The information stored are: storage media, creation and production information, content semantic description, content structural description, low-level content features, information related to the usage of the content, information about the description.

Arslan et. al. [12] describe a video annotation tool used to view, update and delete the semantic data. This data are categorized into five groups: video metadata (video name, length of video, year of production, etc.), object data, event data, subevent data and utility data (audiences, video types, activities, activity roles, etc.).

Moreover, there is the concern of developing tools for wireless environments. Huang [13] proposed a mobile video question/answering system for online annotation and ubiquitous multimedia learning. Ahn [14] developed an embedded multimedia database management system based on MPEG-7 for mobile devices such as PDAs. Sánchez-Nielsen [15] proposed the use of the MPEG-7 standards to produce and publish personalized information according to user’s demand on podcasts environments.
3. Existent Tools

Kosh [1] states that there are three relevant commercial tools for multimedia indexing. First, there is the **IBM Annotation Tool**\(^1\), where each shot in the video sequence can be annotated with many descriptions: static scene, key object, event and other lexicon sets. The descriptions are associated with each video shot and are stored as MPEG-7 descriptions in an XML file. Second, the **Ricoh MovieTool**\(^2\) is a tool that generates MPEG-7 descriptions based on the structure of the video and allows the user to create the structure while watching the video. Third, the **MPEG-7 Audio SpokenContent Tool - Canon**\(^3\) generates a description file from an audio file in the wav format using a speech recognition tool.

Other tools were developed to carry out video indexing process. Some tools such as **Annodex**\(^4\) and **Anvil**\(^5\) do not generate output in standard MPEG-7, but in its own XML format. Other softwares, such as **4M**\(^6\), **MuViNo**\(^7\) and **Vannotea**\(^8\) perform video indexing video using MPEG-7 standards.

The software **4M (MultiMedia Metadata Management)** is a web tool that allows the extraction of characteristics of multimedia objects via MPEG-7, the management of the database (XML) and the treatment of annotations. This tool uses the concept of ontology to describe processes of algorithms necessary for multimedia objects. In **MuViNo** tool, the annotations of the descriptions are taken directly into a tree that represents the location of information, making the registration of complex video segments. In this way, you lose transparency in the operation of the tool, because it does not offer a friendly-user layer to collect user information, but it simply offers the user the structure of the MPEG-7 standards. The **Vannotea** software has an integrated browser allowing the search of multimedia material that can be viewed and recorded in the tool. It also allows other users to view annotations from those who are participating in the session, allows attach files to several segments of video in the annotations and has a space for integration of various types of players.

---

\(^1\) [http://www.alphaworks.ibm.com/tech/videoannex](http://www.alphaworks.ibm.com/tech/videoannex)

\(^2\) [http://www.ricoh.co.jp/src/multimedia/MovieTool/](http://www.ricoh.co.jp/src/multimedia/MovieTool/)

\(^3\) [http://www.cre.canon.co.uk/mpeg7asr/](http://www.cre.canon.co.uk/mpeg7asr/)

\(^4\) [http://www.annodex.net/taxonomy_menu/6/33](http://www.annodex.net/taxonomy_menu/6/33)

\(^5\) [http://www.anvil-software.de/documentation.html](http://www.anvil-software.de/documentation.html)


\(^7\) [http://vitooki.sourceforge.net/components/muvino/](http://vitooki.sourceforge.net/components/muvino/)

4. A Tool for Video Annotation

Based on the features from the available index video tools and works mentioned on Section 2, we describe a set of five basic requirements for developing an indexing video tool for educational environments. First, the software needs to provide means to show audiovisual content stored in a remote repository. Second, it has to allow storing video sites in the remote repository. Third, it must allow marking video segments that contain relevant information in the knowledge field. Fourth, it has to facilitate the annotation of marked video segments with subjective information of high-level semantics. Fifth, it should allow finding, retrieving and displaying of video segments and information associated with them, through a search engine for textual expressions.

An important issue in the development of the tool is portability. The teacher can insert video files in the multimedia repository and someone else can select video segments and perform the necessary annotations. The student can access video files, watch these selected videos (or parts of them) and search some information using keywords. Both activities, on the server and client sides, must be transparent to the user and platform independent. Thus, the application was developed to run locally, within a web portal, or remotely, accessing the repository of videos on the server. To ensure this portability, the tool was developed in Java and was incorporated into the portal as a Java Applet.

4.1 Semantic Video Content Model

Our tool uses the MPEG-7 standard to create video segment annotations, ensuring them the interoperability and some standardization. The MPEG-7 standard defines a set of metadata used to describe some semantic content. In this work, we have adopted only a few of these fields which represents the high level description of the video at a and at the same time are generic enough to be used for any education area. The set of metadata selected is:

1. **Annotation Theme**: represents the label of the annotation. In general, it is a short textual expression used to identify the annotation context;
2. **Annotation Author**: stores the identity of the author who made the description;
3. **Scene local**: is a textual expression that describes the location in the video segment or the location to which that segment refers;
4. **Scene characters**: describes the agents and/or actors in the scene. Individuals who perform relevant actions on the scene;
5. **Scene objects**: identifies some inanimate artifacts present in the scene;
6. **Scene action**: describes the actions performed in the video segment selected;
7. **Scene Observations**: is a textual expression that may contain full details about the video segment.

The first six fields above provide specific descriptions, facilitating the automated search, where as the free text field (Scene Observations) support descriptions in high-level semantics.

The user is not required to know the MPEG-7 standard. Our tool simplifies the annotation process by providing simple and easy-to-understand user interfaces, which enable the user to mark the video segment, through specific controls and menus that supply some metadata form fields. Moreover, another important aspect is to allow the display of content to be described, so the user can view the features to be described at the same time that view the video.

It was also necessary to design a mechanism to locate and recover the segments of videos described, where the user can inform one or more keywords, using the same register structure. By identifying the segment that contains the information sought, the tool temporarily located within the segment of video and displays the user.

### 4.2 Video Annotator Architecture

The tool presented in this paper was implemented on a client-server architecture basis as showed in the Figure 1.

![Client-server architecture for video annotation](image-url)
The process server manages the connections among clients, the video collection storage and their related annotations. This process is also responsible for the search and retrieval of audiovisual content from textual expressions. The client process is the user interface that is connected to the server and supplies the user with the features of the tool. Due to the need to support current main video formats and to support the streaming of these formats, we have decided to integrate the project VideoLan [18] in our tool. A comparative analysis [17] has emphasized that the VLC player is the most suitable tool for the complete transparency and portability for video manipulation systems. The VLC player was integrated with success through the implementation of a Java visual component that encapsulates all the interaction between the JVLC (VLC Java library) and the native VLC player libraries. In our first prototype, we have implemented a protocol for multimedia data transmission, based on TCP/IP. Through this protocol, the complete archive of the video is downloaded from the server, then the video sequence requested by the user is located and displayed.

The client user interface is divided into two distinct areas: one area to content annotation and another one to content search and retrieval activities. This interface is divided into 4 panels (Figure 2).

In the Figure 2, the left panel (area 1) is responsible for the interaction with the server. Through that panel, the user can select either a remote video collection to
locally watch them, or to upload videos to the remote collection. The upper central panel (area 2) incorporates a video player together with all the resources needed for video reproduction selected by the user. With a double click in this area, the video is displayed in full screen, so that the user can devote all his or her attention to the video. The lower central panel (area 3) is intended for creating notes related to the selected video. In this panel, you find the controls needed for choosing a time slice of a video segment and inserting metadata information. Finally, the right panel (area 4) shows a list of previously created annotations related to the selected video and copyright information of audiovisual content. In this panel, the user can select some annotations that perhaps he or she would like to see.

The Figure 3 shows the interface for searching and retrieving of audiovisual content. This tool component is divided into 3 panels.

The lower left panel (area 1) encapsulates a video player, which is used to display annotated video segments (i.e., relevant fragments previously recorded) selected by the user. The lower right panel (area 2) shows the textual annotations for the selected video segment and their copyright information. Finally, the upper panel (area 3) provides the controls and forms for searching the audiovisual content. In this panel, the searching results are displayed in two possible views: list and tree formats. The Figure 4 shows the information under the following notation:

\[<\text{Video name}>; <\text{occurrence}>; <\text{description theme}>\]
This classification was adopted so that the user can easily filter the results according to their relevance.

![Search results](image)

Figure 4: Search results view diagrams

5. Conclusions

Nowadays, there is a huge demand for multimedia technology on the Internet and educational environments. Digital annotation of multimedia objects such as videos is becoming the main activity for collaborations in learning environments. Many of the annotation systems presented in Sections 2 and 3 have been designed for stand-alone environments. We proposed a video tool annotation for a virtual education scope like learning environments or distance learning. The idea is being applied in a larger project which aims to supply teachers and students with a video retrieval tool to be used in classroom contexts.

The tool described in this work is very simple, which has some basic fields for video indexing through semantic content annotations. In the scope of a larger project of learning environments, this first version presents features that can be further studied in order to incorporate more functionalities.

Currently, our annotation tool is still a prototype and has several aspects that we need to improve. We plan to study the inclusion of new fields in the description of the scenes, checking the possibility of using specific terms, for instance the surgical instruments, surgical types and so on, to the areas of education. Furthermore, a study should be performed on the permissions / restrictions of access to users in web portals, and security issues involved.
References