PEANO: Pictorial Enriched ANnotation of VideO

C. Grana, R. Vezzani, D. Bulgarelli, G. Gualdi, R. Cucchiara
Dipartimento di Ingegneria dell'Informazione
University of Modena and Reggio Emilia
Via Vignolese 905/b, 41100 Modena, Italy

M. Bertini, C. Torniai, A. Del Bimbo
Dipartimento di Sistemi e Informatica
University of Florence
Via S.Marta,4 Firenze, Italy

ABSTRACT

In this DEMO, we present a tool set for video digital library management that allows i) structural annotation of edited videos in MPEG-7 by automatically extracting shots and clips; ii) automatic semantic annotation based on perceptual similarity against a taxonomy enriched with pictorial concepts iii) video clip access and hierarchical summarization with stand-alone and web interface iv) access to clips from mobile platform in GPRS-UMTS video-streaming. The tools can be applied in different domain-specific Video Digital Libraries. The main novelty is the possibility to enrich the annotation with pictorial concepts that are added to a textual taxonomy in order to make the automatic annotation process more fast and often effective. The resulting multimedia ontology is described in the MPEG-7 framework. The PEANO (Perceptual Annotation of Video) tool has been tested over video art , sport (Soccer, Olimpic Games 2006, Formula 1) and news clips.

Categories and Subject Descriptors

H3.7 [**Information Storage and Retrieval**]: Digital Libraries – *collection, standards, system issues.*

General Terms

Algorithms, Documentation, Standardization.

Keywords

MPEG-7, video annotation, sports video

1. TOOLS FOR VIDEO ACCESS

The increasing spread of Video Digital Library calls for the design of efficient Video Data Management Systems to manage the video access, provide summarization, similarity search and support queries according with available annotations.

Video summaries, in hierarchical or parametric ways, are necessary to provide very compressed representation of the video without losing crucial contents and to allow efficient browsing as well as a fast overview of the original contents by dropping the time spent on tedious operations such as fast forward and rewind. In this aspect, video summarization is popularly regarded as a good approach to the content-based representation of videos [1].

Video summaries require a structural annotation of videos into elementary parts such as clips or frame sequences extracted possibly in an automatic way according with the content variation. In edited videos, clips are extracted by means of shot detection and other shot partitioning processes. To this aim, we developed a former tool, called **VideoBrowse** (available on web at http://astral.ced.tuc.gr/delos/) to open videos, provide hierarchical summarization at shot and sub-shot level, compare two videos and edit video parts. It was formerly defined for Ferrari s.p.a. to support the analysis of Formula 1 race videos.

Currently, this work is supported by the DELOS Network of Excellence on Digital Libraries, as part of the IST Program of the European Commission (Contract G038-507618 2004-2008) and is carried out in collaboration between a pool of European universities. Videos should be available for different types of access, either downloading or streaming from standard and mobile platforms. As a prototype of the MIMA (Multimedia Interfaces for Mobile Applications) sub-task of DELOS, a tool for mobile access to MPEG-7 video summary is now available: The MOMA (MObile Multimedia Access) tools exploits an UMTS video streaming protocol to selectively access to video clips . It is developed in collaboration with University of Roma La Sapienza too [2].

2. TOOLS FOR VIDEO ANNOTATION

These tools allows a stand alone and Web interface for video clips access: videos can be automatically annotated in a structural manner (describing the structure of shots and sub-shots) and manually or automatically annotated according with a defined taxonomy. The taxonomy contains categories and subcategories: that are described in MPEG-7 and are manually defined with a specific interface. In the future they will be integrated with a specific multimedia ontology creations tools such as Graphonto defined in the DELOS project too [3].
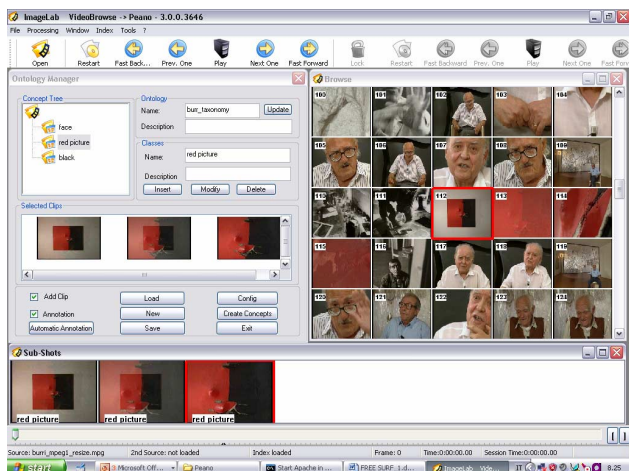
Fig. 1. A screenshot of the developed annotation tool.

Structural annotation is provided with an automatic process that extract shots. The automatic subdivision of videos into clips is a widely faced problem, and several solutions are available. Our system uses a subsystem that defines a linear transition detection process as described in [4] to extract shots. Results of this methods have been tested and presented in TRACVID2005 arena [5] Shots are further divided into sub-shots by means a fuzzy c-means frame clustering. Fig. 1 shows a screenshot of VideoBrowse with shot and sub-shot visualization of clips of a video on the informal artist Burri and his paintings.

The main novelty of this work is the **PEANO** module. The name PEANO (Pictorial Enriched annotation of Video) is in the honor of the famous Italian mathematician Peano that is one of the founders of mathematical logic and the first one that used the name "class" to say prototype or set of elements of the same type. The PEANO tool allows a user friendly fast annotation of clips by means of the classes defined in a textual taxonomy or in an ontology. It can be used to divide large training set of clips in manually defined classes. In video clips, the semantics associated with a class cannot be often represented with a single perceptual aspect but normally many clips of the same class have a different pictorial aspects [6]. This is well known, and automatic annotation by similarity is hard to be performed against a single prototype for each class. Many approaches use many examples for each class. The PEANO module uses a hierarchical clustering method, based on *Complete Link* [6], to reduce the number of clip of each class, in a variable number of prototypes (or pictorial concepts) keeping only some representatives, which capture the most significant aspects of a set of clips. In Fig. 1 for instance in the left-most window, some

prototypes of the class "red picture" are indicated. This clustering method allows an average reduction of about 80-90% of examples in the training set. The elicited set of pictorial concepts are used in the automatic annotation process. A new video is automatically divided into clips; each clip is compared with the set of prototypes with a nearest neighbor classifier in a feature space composed by several perceptual features: Scalable HSV histograms, YCbCr color layout, and parametric motion using MPEG motion vectors. The features and the resulting annotation are saved in an MPEG-7 description.

The annotation allows for a fast retrieval of similar clips: given a clip of a video, the other clips that have been annotated according with the same pictorial concept or the same textual class are ranked. Results of similarity search are promising in term of recall and strongly depend on the Video Digital Library domain. Whenever the videos contain a perceptual regularity, such as in many sport videos, an average recall of about 70% was reached in our experiments.

ACKNOWLEDGMENTS

REFERENCES

[1] M. Yeung, B.L. Yeo, *Segmentation of video by clustering and graph analysis*, in Computer Vision and Image Understanding, Vol. 71 No. 1, July, pp 94-109, 1998.

[2] E. Bertini, G. Santucci and T. Catarci, " MIMA: Multimedia Interfaces for Mobile Applications" Proc. Of 2006 Second International Symposium on Communications, Control and Signal Processing. Marrakesh 2006

[3] C. Tsinaraki, P. Polydoros and S. Christodoulakis "GraphOnto: A Component and a User Interface for the Definition and Use of Ontologies in Multimedia Information Systems" Proc. Of AVIVDILB Padua, I,2005

[4] C. Grana, G. Tardini, and R. Cucchiara, MPEG-7 Compliant Shot Detection in Sport Videos, in Proceedings of the IEEE International Symposium on Multimedia (ISM 2005), Irvine (CA), USA, (2005) 395-402.

[5] http://www-nlpir.nist.gov/projects/tvpubs/tv.pubs.org.html

[6] M. Bertini, R. Cucchiara, A. Del Bimbo, and C. Torniai, Video Annotation with Pictorially Enriched Ontologies, Proceedings of IEEE International Conference on Multimedia and Expo, Amsterdam, The Netherlands, (2005) 1428-1431.

[7] A.K. Jain, and R.C. Dubes, Algorithms for clustering data, Prentice-Hall, Englewood Cliffs, NJ (1988).