Arneb: a Rich Internet Application for Ground Truth Annotation of Videos

Thomas Alisi, Marco Bertini, Gianpaolo D'Amico, Alberto Del Bimbo,
Andrea Ferracani, Federico Pernici and Giuseppe Serra
Media Integration and Communication Center, University of Florence, Italy
{alisi, bertini, damico, delbimbo, ferracani, pernici, serra}@dsi.unifi.it
http://www.micc.unifi.it/vim

ABSTRACT

In this technical demonstration we show the current version of Arneb[1], a web-based system for manual annotation of videos, developed within the EU VidiVideo project. This tool has been developed with the aim of creating ground truth annotations, that can be used for training and evaluating automatic video annotation systems. Annotations can be exported to MPEG-7 and OWL ontologies. The system has been developed according to the Rich Internet Application paradigm, allowing collaborative web-based annotation.

Categories and Subject Descriptors

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*Video*; H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*Web-based services*

General Terms

Algorithms, Experimentation

Keywords

Video annotation, ground truth, video streaming

1. INTRODUCTION

The goal of the EU VidiVideo project is to boost the performance of video search by forming a 1000 element thesaurus of automatic detectors to classify instances of audio, visual or mixed-media concepts. In order to train all these classifiers [1], using supervised machine learning techniques such as SVMs, there is need to easily create ground-truth annotations of videos, indicating the objects and events of interest. To this end a web based system that assists in the creation of large amounts of frame precise manual annotations has been created. Through manual inspection of the broadcast videos, several human annotators can mark up the start and end time of each concept appearance, adding frame accurate annotations.

Web-based annotations tools are becoming very popular because of some mainstream video portals or web televisions like YouTube, Splashcast and Viddler. In fact, online users now are able to insert notes, speech bubbles or simple subtitles into the videos timelines with the purpose of adding further informations to the content, linking to webpages, advertising brands and products. These new metadata are managed in a collaborative web 2.0 style, since annotations can be inserted not only by content owners (video professionals, archivists, etc.), but also by end users, providing an enhanced knowledge representation of multimedia content.

Our system starts from this idea of collaborative annotations management and takes it to a new level of efficacy. All the previous systems, in fact, use annotations composed of non structured data and do not use any standard format to store or export it. All these metadata (tags, keywords or simple sentences) are not structured and do not allow to represent composite concepts or complex information of the video content. Our solution proposes a semantic approach, in which structured annotations are created using ontologies, that permit to represent in a formal and machine-understable way a video domain. The import and export functionalities of both annotations and ontologies, using MPEG-7 standard and OWL ontologies, allow an effective interoperability of the system. The system has been intensively used to annotate several hours of videos, creating more than 25,000 annotations, by B&G (Dutch national audiovisual institute) professional archivists.

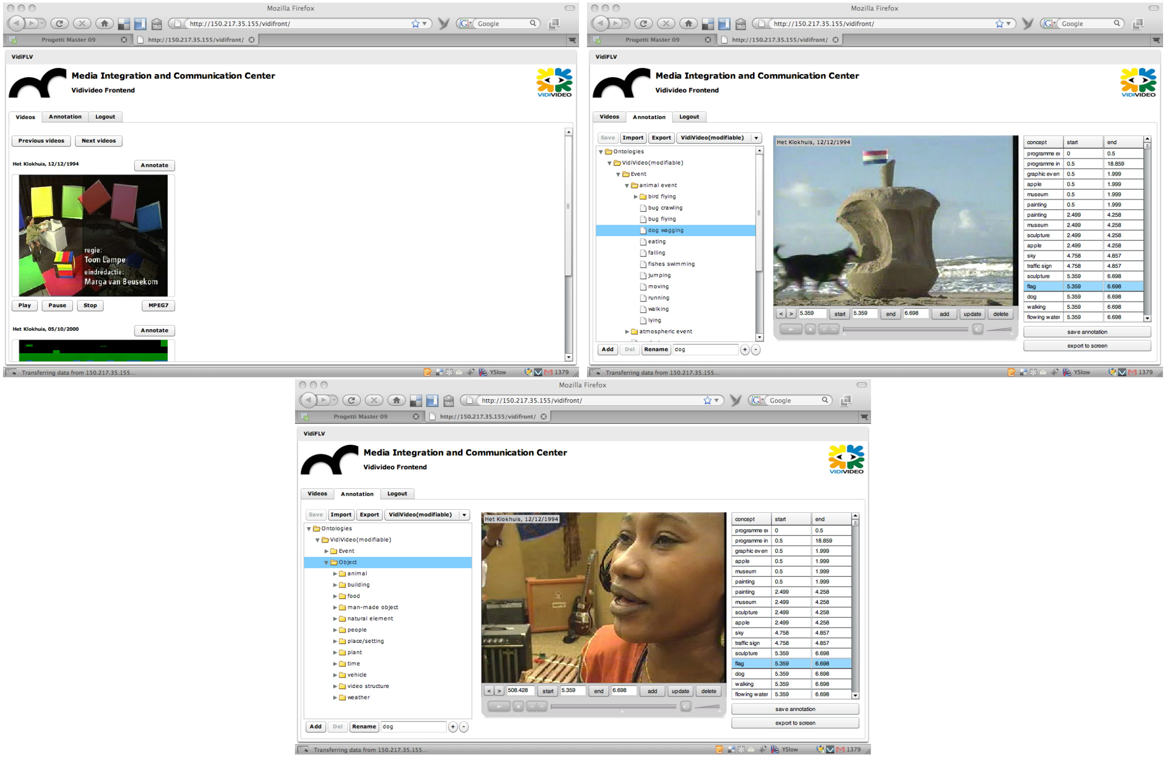
2. THE SYSTEM

The video annotation system allows to browse videos, select concepts from an ontology structure, insert annotations and download previously saved annotations. The system can work with multiple ontologies, to accomodate different video domains. Ontologies can be imported and exported using the MPEG-7 or OWL ontology, or created from scratch by users. The system administrator can set the ontologies as modifiable by end-users. Annotations, stored on a SQL database server, can be exported to both MPEG-7 and OWL ontology formats. The use of MPEG-7 provides interoperability of the system, but this format reflects the "structural" lack of semantic expressiveness, that can be obtained by its translation into an ontology language [2]. To cope with this problem has been developed the VidiVideo DOME (Dynamic Ontology with Multimedia Elements)[2] ontology, mod-

[1] Arneb is a hare, one of the preys of the mythical hunter Orion. It is chased by Orion's hound, Sirio.

[2] http://www.micc.unifi.it/dome/

MM'09, October 19–24, 2009, Beijing, China.



concept	start	end
programme e	0	0.5
programme in	0.5	18.859
graphic even	0.5	1.999
apple	0.5	1.999
museum	0.5	1.999
painting	0.5	1.999
painting	2.499	4.258
museum	2.499	4.258
sculpture	2.499	4.258
apple	2.499	4.258
sky	4.758	4.857
traffic sign	4.758	4.857
sculpture	5.359	6.698
flag	5.359	6.698
dog	5.359	6.698
walking	5.359	6.698
flowing water	5.359	6.698

concept	start	end
programme e	0	0.5
programme in	0.5	18.859
graphic even	0.5	1.999
apple	0.5	1.999
museum	0.5	1.999
painting	0.5	1.999
painting	2.499	4.258
museum	2.499	4.258
sculpture	2.499	4.258
apple	2.499	4.258
sky	4.758	4.857
traffic sign	4.758	4.857
sculpture	5.359	6.698
flag	5.359	6.698
dog	5.359	6.698
walking	5.359	6.698
flowing water	5.359	6.698

Figure 1: Annotation interface (in clockwise order): browsing videos, searching a concept, annotating.

elled following the Dynamic Pictorially Enriched Ontology model [3]. DOME defines and describes general structural video concepts and multimedia concepts, in particular dynamic concepts. To cover different fields, videosurveillance, news and cultural heritage, DOME is composed by a common core and by three domain ontologies developed for the video domains used within the VidiVideo project: DOME Videosurveillance, DOME News, DOME Cultural Heritage.

The system is implemented according to the Rich Internet Application paradigm: the user interface runs in a Flash virtual machine inside a web browser. RIAs offer many advantages to the user experience, because users benefit from the best of web and desktop applications paradigms. Both high levels of interaction and collaboration (obtainable by a web application) and robustness in multimedia environments (typical of a desktop application) are guaranteed. With this solution installation is not required, since the application is updated on the server, and run anywhere regardless of what operating system is used. The user interface is written in the Action Script 3.0 programming language using Adobe Flex and Flash CS4.

The backend is developed in the PHP 5 scripting language, while data are stored in a MySQL 5 database server. All the data between client and server are exchanged in XML format using HTTP POST calls. The system is currently based both on open source tools (Apache web server and Red 5 video streaming server) or freely available commercial ones (Adobe Flash Media Server free developer edition). All videos are encoded in the FLV format, using H.264 or On2 VP6 video codecs, and are delivered with the Real Time Messaging Protocol (RTMP) for video streaming.

3. DEMONSTRATION

We demonstrate the creation of ground-truth annotations of videos using a tool that allows a collaborative process through an internet application. Differently from other web based tools the proposed system allows to annotate the presence of objects and events in frame accurate time intervals, instead of annotating a few keyframes or having to rely on offline video segmentation. The possibility for the annotators to extend the proposed ontologies allow expert users to better represent the video domains, as the vidos are inspected. The import and export capabilities using MPEG-7 standard format allow interoperability of the system.

Acknowledgments.
This work is partially supported by the EU IST VidiVideo project (www.vidivideo.info - contract FP6-045547).

4. REFERENCES

[1] Cees G. M. Snoek, et al. The MediaMill TRECVID 2008 Semantic Video Search Engine. In *Proc. of 6th TRECVID Workshop*, Gaithersburg, USA, November 2008

[2] N. Simou, V. Tzouvaras, Y. Avrithis, G. Stamou, and S. Kollias. A visual descriptor ontology for multimedia reasoning. In *Proc. of Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2005)*, Montreux (Switzerland), April 2005.

[3] M. Bertini, R. Cucchiara, A. Del Bimbo, C. Grana, G. Serra, C. Torniai and R. Vezzani. Dynamic Pictorially Enriched Ontologies for Video Digital Libraries. In *IEEE Multimedia*, to appear, 2009.