Video Analysis Tools for Annotating User-Generated Content from Social Events

Rodrigo Laiola Guimarães[1], Rene Kaiser[2],
Albert Hofmann[2], Pablo Cesar[1] and Dick Bulterman[1]

[1] CWI: Centrum Wiskunde & Informatica
Science Park 123, 1098 XG Amsterdam, The Netherlands
{rlaiola, p.s.cesar, dick.bulterman}@cwi.nl

[2] Institute of Information and Communication Technologies
JOANNEUM RESEARCH
Graz, Austria
firstname.lastname@joanneum.at

Abstract

In this demo we present how low-level metadata extraction tools have been applied in the context of a pan-European project called Together Anywhere, Together Anytime[1] (TA2). The TA2 project studies new forms of computer-mediated social communications between spatially and temporally distant people. In particular, we concentrate on automatic video analysis tools in an asynchronous community-based video sharing environment called MyVideos, in which users can experience and share personalized music concert videos within their social group.

In the MyVideos scenario, people attend to an event such as a school concert rehearsal and take videos, but not everybody films all the time. The purpose of each video is primarily personal: each family is interested in capturing their own child, but also enough context information from the concert to provide some background. Each family is interested in creating a video fragment for the personal family archives, or short clips that can be sent to family members who were not able to be at the show.

After the concert has taken place users can upload their video clips to the MyVideos platform. Before making such material available through a Web-based application, we use the Semantic Video Annotation Suite[2] (SVAS) and a video alignment tool [1] to automatically extract the metadata necessary to organize and tag the content subsequently. Besides generating potential key frames, the SVAS tool (see Fig. 1) detects severe unusual material that has the same appearance as shot boundaries – like very rapid camera movements, heavy unsteadiness or people crossing the picture close to the camera – and generates a set of annotations using the MPEG-7 standard[3], that then is stored in the MyVideos database. Other functionalities such as person and instrument recognition is currently under development –

---

[1] http://www.ta2-project.eu/

[2] http://www.joanneum.at/en/digital/avm/products-solutions-services/semantic-video-annotation.html

[3] http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm

annotating user-generated content is challenging because the video encoding, the quality, and the lighting are not always optimal. Since users record, at their will, different parts of the concert, a key phase of the content preparation process is the alignment of all recorded video clips to a common timeline. For that matters, in addition to SVAS, a video alignment tool has been developed in the context of TA2 [1], and it can accurately align temporally user-generated video clips based on a high-quality audio stream recorded throughout the event. The results of such tools annotate the video clips and provide the means for easy search and navigation on the MyVideos front-end (see Fig. 1).

In our demo we present the MyVideos Web application (using an iPad) and the video analysis toolset (in the laptop), in their current, though not final, state of development. We give special attention to the automatic extraction processes of the SVAS tool, and we show how the metadata obtained allows for easy exploration of the concert media space (e.g., clip suggestion with relevant fragments within a clip for the user to watch or to share) and review of the video clips participants have recorded, along with their annotations.
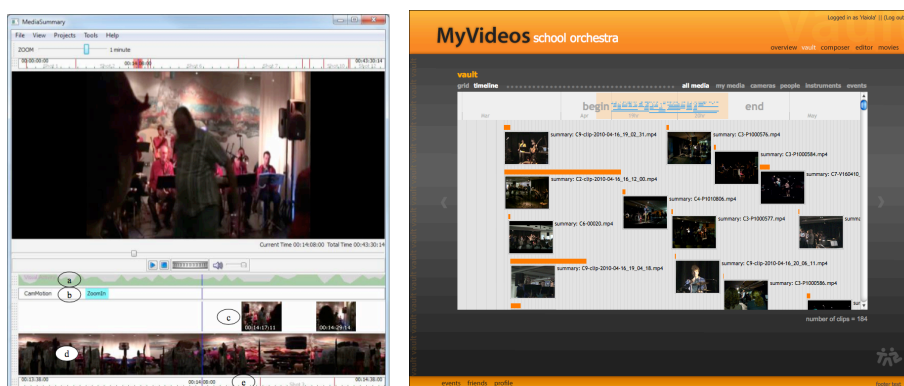

Fig. 1. Semantic Video Annotation Tool (on the left). MyVideos Web Application: Timeline View (on the right).

References

1. Korchagin, D., Garner, P. N., and Dines, J.: Automatic Temporal Alignment of AV Data with Confidence Estimation. In: IEEE Proceedings of the International Conference on Acoustics, Speech and Signal Processing (2010)