Fast Annotation of Video Objects for Interactive TV

Helmut Neuschmied
Joanneum Research
Steyrergasse 17-19
8010 Graz
helmut.neuschmied@joanneum.at

Remi Trichet
Institut EURECOM
06904 Sophia Antipolis
France
trichet@eurecom.fr

Bernard Merialdo
Institut EURECOM
06904 Sophia Antipolis
France
merialdo@eurecom.fr

ABSTRACT
In this demonstration, we present the Annotation Tool that is being developed in the porTiVity project to annotate video objects for Interactive Television programs. This tool includes various video processing components to structure and speed-up the annotation process, such as shot segmentation, key-frame extraction, object tracking and object redetection. A specific feature is that the tool includes a preprocessing phase where a quantity of information is precomputed, so that the annotation itself can be done quite rapidly.

Categories & Subject Descriptors: D.2.2 Design Tools and Techniques, Visual C++

General Terms: Design

Keywords:
video annotation tool, generic object tracker, object re-detection, real-time application

1. INTRODUCTION
Nowadays, the overload of information is one of the most important issues in the domain of information retrieval. As the different methods for exploring and describing data grow slower than the amount of information, it becomes cumbersome for the user to find relevant information.

In this paper, we present the *porTiVity Annotation Tool* designed to assist authors in annotating video sequences. Each video object identified by the user in an frame is automatically localized in the entire shot with bounding boxes and can be linked with media information such as text, images, internet address, or other videos. This tool can also perform image segmentation, shot segmentation, and the redetection of the user identified object in other shots.

One of the particularities of this tool is the presence of a separated preprocessing component where most of computational work needed during the annotation phase could be done in advance. This allows the annotation operations to be very fast, responding to the requirements of an ergonomic annotation tool.

The rest of this article will be decomposed as follows: First of all, we will briefly describe the porTiVity project. In the second part, the overall structure of the Annotation Tool will be schemed. The third and the fourth sections will respectively present the two distinct components of the tool: the *MediaAnalyze Tool* and the *Annotation Tool* itself. Finally, the last part will show some advanced features.

MM'07, September 23–28, 2007, Augsburg, Bavaria, Germany.
ACM 978-1-59593-701-8/07/0009.

2. THE PORTIVITY PROJECT
Our work takes place in the context of the European Portivity project [1].This project for interactive television aims to realize direct interactivity with moving objects on hand-held receivers. In this system, the video producer will annotate the video program by defining video objects in the video sequence, and attaching to them some additional content (for example, text, images, videos, web reference, etc…). On the receiver side, the user watching the video program will be able to select the active video objects and immediately access the corresponding additional content.

3. SYSTEM STRUCTURE
The porTiVity annotation architecture can be seen in Figure 1. A pre-processing tool, called MediaAnalyze Tool, performs most of the computational work of video analysis, and stores the results in a MPEG-7 database. With the Annotation Tool, the user selects video objects and relates them to semantic descriptions. In order to reduce the annotation effort, the Annotation Tool contains automatic object tracking and object redetection functionalities, which uses the preprocessed data to reduce the computation time. All produced metadata is stored in the ISO standard MPEG-7 [2].

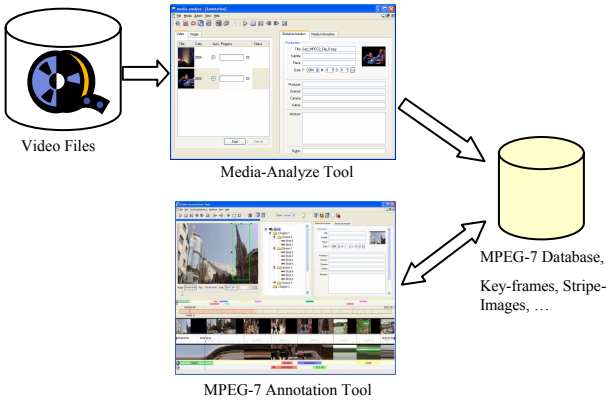

Figure 1: Annotation system

4. PREPROCESSING TOOL
The MediaAnalyze Tool is a separate application. For content management purposes, the tool provides the possibility to define certain global attributes for the video, for e.g. the title and some production information. The rest of the process is entirely automatic. During the automatic content analysis, metadata is extracted for video navigation and structuring like shots and key-frames, and image keypoints and their associated descriptors required for object tracking and object re-detection are calculated.

The analysis process can be started for several videos and then the process can run over night.

The core part of this program is a module framework. The framework serves as an execution environment for analysis modules. The modules are interconnected by the framework to constitute a so-called module graph (see **Figure 2**). The module graph is defined by a XML file.

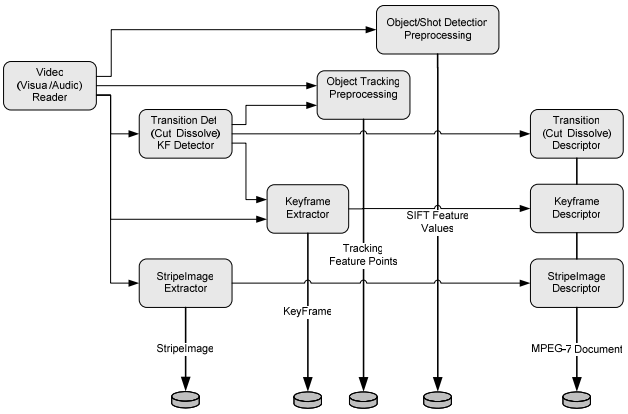

Figure 2: Analyses module graph of the MediaAnalyze Tool

5. ANNOTATION TOOL

As soon as the automatic analysis of a video is completed, the produced metadata description can be displayed, edited and extended by the MPEG-7 Annotation Tool (see Figure 3). This tool has a number of views which enable fast and easy navigation in the video. Through the key frames and the stripe image, one gets a quick overview of the video content. There are two time lines, one for the whole video time and one which shows only a selected time period (time zoom). In the time lines, the shot boundaries and the dissolves are displayed and they can also be edited. There is also the possibility to structure the video depending on the video content. For example shots can be grouped to scenes; scenes can be combined to build chapters and so on. This structure yields a kind of table of contents and is displayed by a separate view. Depending on the selected structural element different textual annotations are possible. These are for example the title of the structural element, content description, remarks, and specifications about time, location, and persons.

For object annotation the integrated video player has drawing functionalities. Objects can be specified by drawing a rectangle or a polygon around them. Once an object location is specified (see green rectangle in Figure 3) it is possible to start an automatic search for a similar region in other shots of the video. The result of the object search is displayed in a separate key-frame view. In each of the displayed key-frames the found object location is drawn in. The search result can be edit manually. After that, the location of the found objects can be tracked in each shot using a keypoint based generic object tracker [3]. The object redetection and tracking functionalities are very fast because they use the preprocessed visual feature values.

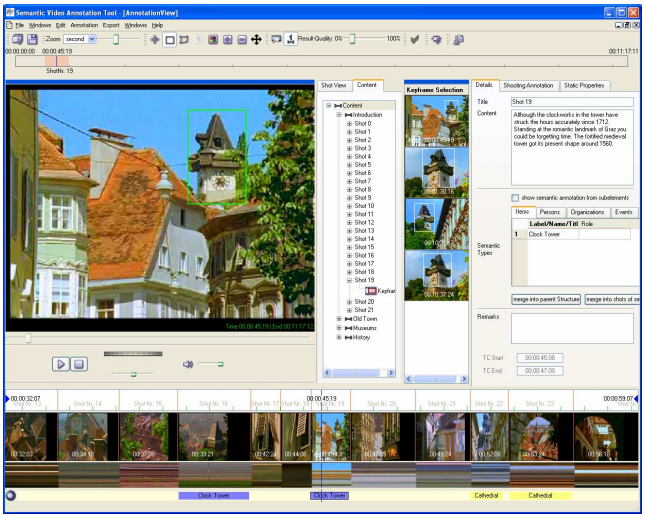

Figure 3: User interface of the MPEG-7 Video Annotation Tool

6. VIDEO ANNOTATION FOR INTERACTIV TV

In the proTiVity system the produced MPEG-7 annotation will be used to produce the interactive content for portable and mobile devices. A tool is under development for generating MPEG-4 LASeR [4] applications. It will use the object description and video structure information. Interactivity can be provided by inserting of predefined menus and buttons in the video and by assigning additional content to the annotated objects. Additional content can be image, video or audio data, HTML pages and LASeR scenes. It can also be a LASeR script which enables to bring in some functionality like personalization of the additional content or a game logic.

7. ACKNOWLEDGMENTS

Part of this work has been supported by the European Commission through the IST-project porTiVity.

8. REFERENCES

[1] Gerhard Stoll et all, *porTiVity: new Rich Media iTV Services for handheld TV*, 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies, London, 1st December 2005

[2] MPEG-7: ISO/IEC 15938:2001.

[3] Trichet, Rémi;Mérialdo, Bernard, *Generic object tracking for fast video annotation*, VISAPP 2007, 2nd International Conference on Computer Vision Theory and Applications, 8 - 11 March, 2007 Barcelona, Spain

[4] MPEG-4 LASeR: ISO/IEC 14496-20