

# Generating TV Summaries for CE-devices

Gerhard Mekenkamp,  
Mauro Barbieri  
Philips Research  
Prof. Holstlaan 4, 5656 AA,  
Eindhoven, The Netherlands  
+31 40 27 44491  
{gerhard.mekenkamp,  
mauro.barbieri}@philips.com

Benoit Huet  
Itheri Yahiaoui  
Bernard Merialdo  
Institut Eurécom  
06904 Sophia-Antipolis,  
France  
+33 0493002679  
{huet, yahiaoui,  
merialdo}@eurecom.fr

Riccardo Leonardi  
Università degli Studi di  
Brescia  
Piazza Mercato 15,  
25121 Brescia,  
Italy  
+39 030 3715434  
leon@ing.unibs.it

Michael Rose  
Technical University of  
Denmark  
Anker Enggelundsvej 1  
DK 2800, Kgs Lyngby,  
Denmark  
+45 4525 5172  
rose@cti.dtu.dk

## ABSTRACT

Automatically generated summaries of TV content are indispensable for content selection and navigation in CE-devices. We show two different types of summaries: Short video trailers and visual overviews consisting of representative frames. The demo does not only show the feasibility of the proposed algorithms, but also shows how different types of generated summaries can be used in future CE-devices.

## Categories and Subject Descriptors

H.3.1 [Information Storage and Retrieval]: Content Analysis and Indexing – *abstracting methods*.

## General Terms

Algorithms, Experimentation.

## Keywords

Content analysis, video summaries.

## 1. INTRODUCTION

The amount of digital content stored in the house is rapidly increasing. Besides storage on CDs and DVDs, hard disks are now finding their way in consumer products. Hard disk based personal video recorders will change the way we watch television mainly due to the option to pause live TV by simultaneously recording and playing back the same programme. Easy recording functions will make sure that there is always something nice to watch. Automatic recording agents can select programmes for recording based on a user profile. This way the amount of programmes that are stored on the disk can be quite large. A 100Gbyte hard disk can already store some 50 hours of TV programmes in broadcast quality (4-5 Mbits/sec). With the hard disk capacity doubling every year and the introduction of new efficient video compression algorithms soon hundreds of hours of video can be stored. This requires advanced functionality for browsing and navigating through video content, making it easier

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Multimedia'02, December 1-6, 2002, Juan-les-Pins, France.

Copyright 2002 ACM 1-58113-620-X/02/0012...\$5.00.

to find specific information or to help decide what to watch.

The amount of available processing power is large enough to justify the use of content analysis techniques. In the Spation (Services Platforms and Applications for Transparent Information Management in an in-home Network) project [1] we apply content analysis techniques to allow easy access to the large amount of digital content that will be stored in the house. Besides large storage capacities in CE-devices we also expect devices to become interconnected. Using handheld devices that connect wirelessly to various devices in the house it will be possible to search, retrieve and navigate through content stored on many devices in the home.

In this demo we focus on the automatic generation of video summaries using content analysis techniques and show its application in a home system. A Personal Video Recorder analyses recorded TV programmes and generates summaries. A handheld device is used to wirelessly browse through the content stored on the PVR's hard disk and select generated summaries.

## 2. TYPES OF GENERATED SUMMARIES

The large storage capacity available in future personal video recorders creates the need for automatically generated summaries. We distinguish between two different approaches in creating summaries.

The first is a video trailer. While a summary should capture the essence of the content as well as possible, a trailer is used to help the user decide to watch the TV programme or not. A trailer is generated by analysing the content and selecting clips focusing on entertainment rather than accuracy. These clips are concatenated and played as a video trailer. This way the user can quickly see what interesting programmes are available while still staying with the passive lean backward mode of watching TV.

A second type of summary that we generate represents the content by selecting representative frames and displaying a number of downscaled versions of these frames simultaneously. Using such an overview the user can quickly see what the content is about. When multiple screens with images are available the user can flip through these screens at his own rate. Thus providing an efficient way to check whether it is worth watching the programme, or to decide whether or not he saw the programme already.

In our demonstration we will show both types of trailers allowing to compare the approaches.

### 3. VIDEO SUMMARY APPLICATIONS

Generated summaries can help the user to decide what to watch, to decide whether or not to delete a programme, to find a certain scene, to remember having watched the programme before, or to decide downloading it. These features are indispensable for hard disk based personal video recorders capable of storing hundreds of hours of video. The hard disk can also serve as a cache to browse through a personal video collection stored on DVDs. That way it is possible to watch summaries of DVDs and decide which one to watch. When summaries are stored on the hard disk, browsing can be fast and easy. When a nice programme is found the corresponding DVD can be inserted.

When watching a number of summaries the user should be allowed to stop playback of a summary and see the next one. Often it is possible to decide not to watch a programme in just a few seconds, similar to zapping through TV channels.

Instead of watching the generated summaries on TV it is possible to download the summaries to a handheld device. This way it is e.g. possible to view summaries while travelling. Preferable the length of the generated summaries is under user control. Perhaps to kill some time during travelling the user would prefer somewhat longer trailers. Durations of 30 seconds to 3 minutes seem to be a reasonable range. For picture-based summaries the number of selected representative frames can be set higher.

### 4. VIDEO TRAILERS

To be able to implement automatic generation of trailers in CE devices the algorithms should perform the necessary content analysis in real-time and should be content independent.

During recording low-level content descriptors are extracted from the content. This includes parsing the video stream using scene change detection and computing simple statistics such as the bit rate, the amount of motion, the amount of detail in a scene, the audio volume etc. In [2] we have shown that low-level content descriptors can be extracted in real time using existing MPEG2 encoder chips.

Statistical information of the entire TV programme are analysed after recording. The selection of an action scene, for example, can then be based on the number of detected action scenes in the entire programme. Since the majority of the analysis work was done during recording, the selection of clips by analysing the statistical data requires little time.

We propose to use many low-level statistical features to reduce the dependency on the type of content. An additional advantage is that the more different features are extracted, the lower the needed accuracy of the extraction algorithm of each feature. For some cases additional event detectors can be helpful. For example, detecting an explosion, a car chase, a goal in a soccer match, laughter, applause etc. However, the number of special events to detect can become quite large and the event might heavily depend on the type of programme. Problems with such an approach in a product are clear. Consider running an algorithm that was designed to detect goals in a soccer match on a tennis match or a movie by mistake.

In the demo we show the results of an algorithm that uses low-level statistical data to construct short trailers from arbitrary TV programmes.

### 5. PICTORIAL SUMMARIES

A second type of video summaries consists of pictorial overviews created by selecting representative frames [3] [4]. During the recording, a frame-to-frame difference is computed based on region colour distribution. When the difference between two images is large enough it is selected as a candidate for further analysis. Selected key-frames take part in further analysis, which consists of a clustering approach that is performed after the recording. Images are clustered based on their colour distribution similarity.

Our objective is to build a summary, which allows the identification of the original video after viewing only a randomly selected excerpt of the video. In order to satisfy this maximum recall condition, the construction process is one where both the frequency and location of occurrence of the frame are employed simultaneously to identify the optimal frame. This way representative frames are selected in a way that captures as much information as possible from the original video.

### 6. DESCRIPTION OF THE DEMO

The demo shows a personal video recorder capable of generating video summaries from MPEG2 digital video broadcast streams. Using a handheld device two different types of summaries can be selected. The first type shows a generated summary in the form of a short video trailer. The second type shows a selection of the programme's most representative frames. Furthermore, the summaries can be viewed or stored on the handheld device, as well as displayed on a TV screen. This demo will show the user interface issues involved in viewing summaries on PDA's as opposed to desktop PC's or TV's.

This demo does not only show the feasibility of the proposed algorithms, but also shows how different types of generated summaries can be used in future CE-devices.

### 7. ACKNOWLEDGEMENTS

This work was conducted as part of the Spation project (Services, platforms and Applications for transparent information management in an in-home Network). Spation sponsored by the European Union under the IST program. We would like to thank the partners in the project, Philips Research, Eurécom, University of Brescia, University of Denmark and Tomorrow Focus.

### 8. REFERENCES

- [1] <http://www.extra.research.philips.com/euprojects/spation>, "Home page of the SPATION project".
- [2] N. Dimitrova, S. Jeannin, J. Nesvadba, T. McGee, L. Agnihotri, G. Mekenkamp, "Real time commercial detection using MPEG features", 9<sup>th</sup> Int. Conf. On Information Processing and Management of Uncertainty in Knowledge-Based Systems, July 2002, Annecy, France.
- [3] Ithery Yahiaoui, Bernard Merialdo et Benoit Huet, "Automatic Summarization of Multi-episode Videos with the Simulated User Principle", Workshop on Multimedia Signal Processing, October 2001, Cannes, France.
- [4] Ithery Yahiaoui, Bernard Merialdo et Benoit Huet, "Generating Summaries of Multi-episode Video", IEEE Int. Conf. on Multimedia and Expo, ICME 2002, August 2001, Tokyo, Japan.