MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Architectures for Video Summarization Services Over Home Networks and the Internet

Sam Shipman, Regunathan Radhakrishan, Ajay Divakaran

TR2006-029 January 2006

Abstract

With increasing number of personal digital video recorders, the idea of having summarization fundtionality to view the recorded content has also been gaining importance. The summarization functionality can either be built into the video recording device or can reside in a computing device on the network. In this paper, we present two architectures of the latter case. The first architecture is to enable summarization services over a Home Network. The second architecture is to enable summarization services over the internet. We also present a client-server implementation for video summarization.

IEEE International Conference on Consumer electronics (ICCE)

ARCHITECTURES FOR VIDEO SUMMARIZATION SERVICES OVER HOME NETWORKS AND THE INTERNET

Sam Shipman, Regunathan Radhakrishan and Ajay Divakaran

Mitsubishi Electric Research Labs, Cambridge, MA, USA.
{shipman,regu,ajayd}@merl.com

ABSTRACT

With increasing number of personal digital video recorders, the idea of having summarization functionality to view the recorded content has also been gaining importance. The summarization functionality can either be built into the video recording device or can reside in a computing device on the network. In this paper, we present two architectures of the latter case. The first architecture is to enable summarization services over a Home Network. The second architecture is to enable summarization services over the internet. We also present a client-server implementation for video summarization.

1. INTRODUCTION

Recently, there has been an increase in the number of digital video recording devices like TiVo that can store several hours video. Also, summarization functionality is being built into these devices to enable efficient access to "interesting" segments of the stored video quickly. With the availability of computational resource in the form of (Personal Computers)PCs and laptops and wireless networking capability (802.11b/g, UWB) at home, one can also imagine having the summarization functionality being offloaded to the computer instead of it being included in the video recording device.

Let us first look at a typical personal video recorder system with built-in summarization function. Figure 1 shows the architecture of a summarization enabled Digital Video Recorder. Here the analysis is performed while recording the video using compressed domain feature extraction and an audio classification framework.

The video and audio signals from a broadcast video are encoded using MPEG2 and AC-3, packetized, and stored onto a disc such as HDD, DVD, Blu-ray medium via buffer. The Audio Classification block in the audio DSP classifies each audio segment as one of several classes (e.g., Applause, Cheering, Excited Speech, Normal Speech, Music, etc.) using low-complexity Gaussian Mixture Models (GMM). The Importance Calculation block calculates the percentage of the significant audio class in a time window of an

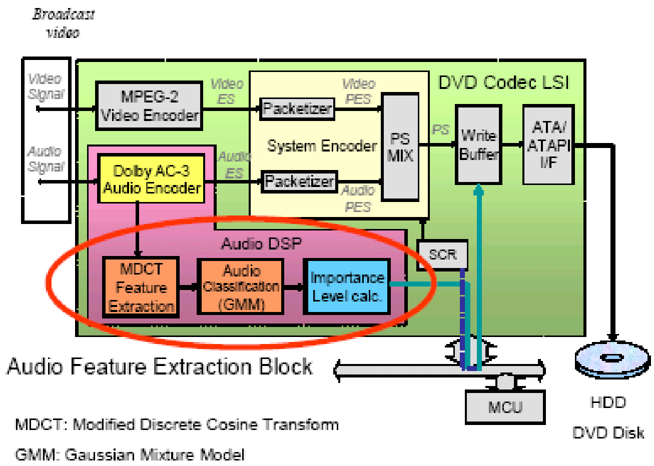

Figure 1: Architecture of a Summarization-Enabled DVD recorder

audio classification data stream to get the importance level for each audio segment. For instance, in sports the significant audio class turns out to be the audience & commentator reaction classes such as applause, cheering and excited speech. From the user-interface point of view, a user can set a threshold on the importance for a desired summary length. The portions with importance greater than the chosen threshold are identified as the highlights. So, skipping to the start position of the highlight scene manually and skipping and playing back only the highlights are functions that are supported. These unique functions give the user a powerful and useful way to browse the large volume content.

In this paper, we propose two architectures for offloading the aforementioned analysis for the summarization functionality. The first architecture is for enabling summarization service over a home network. The second architecture is for enabling summarization service over the internet. We also present a client-server implementation of summarization service.

The rest of the paper is organized as follows. In section 2, we describe the proposed architecture for home networks. In section 3, we describe the proposed architecture for en-

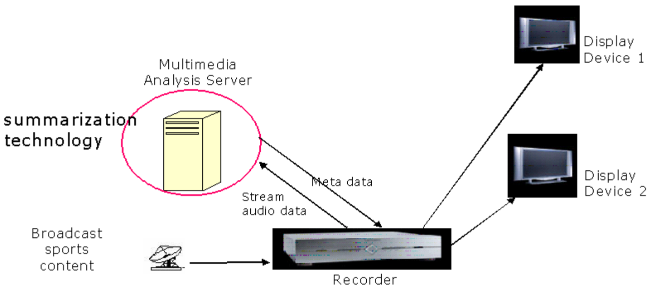

Figure 2: Architecture of a Summarization-Enabled Home Network

hanced multimedia services over the internet. In section 4, we describe a client-server implementation for summarization. In section 5, we conclude with some discussions.

2. ARCHITECTURE FOR SUMMARIZATION SERVICES OVER HOME NETWORKS

In the recent years, wireless access technologies at home such as 802.11 and the Ultra Wide Band (UWB) have been increasing the usable bandwidth at home and thus can enable multimedia home networking applications. Figure 2 shows the architecture of a summarization enabled Multimedia Home Network Architecture. Here the content analysis is performed on a set-top box or a multimedia home server after recording the program onto a HDD. Then, the content bitstream to be analyzed is streamed to the analysis server that has the content analysis technologies. The results of summarization (meta-data for the content) are sent back to be viewed on one or more display devices at home.

3. ARCHITECTURE FOR SERVICES OVER THE INTERNET

In the previous case, we considered the content analysis server to be a part of the home network. However, if the content provider can employ one of these servers to create meta-data for summaries one can imagine delivering only the interesting parts of the content to the end-user. For instance, using mobile phones with video capability one can request a summary to the network service provider for a recent highlight scene of a live game. The request gets forwarded to the content analysis server and the selected highlight segment can be transcoded to adapt to the display size of the mobile phone and delivered. Figure 3 shows the architecture of a network supporting the creation and the distribution of summarized content.

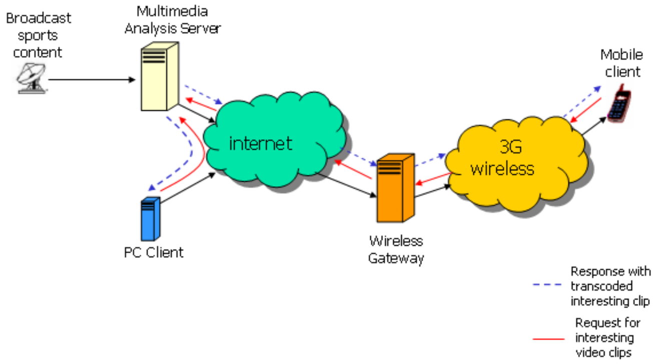

Figure 3: Architecture for Summarization-Enhanced Multimedia Services over the Internet

4. A CLIENT-SERVER IMPLEMENTATION FOR SUMMARIZATION

In this section, we describe an example client-server implementation for summarization service. The client in this example implementation is a digital video recorder and the server is a home PC. Both the client and the server are connected to a LAN or a wireless LAN. Now let us look at how audio classification based analysis for summarization can be accomplished by streaming an ac3 bitstream from the client to the server.

The server constantly listens for a socket connection request at a specific port. The client has access to complete ac3 file recorded onto local storage. It first requests a socket connection from the server. After the connection has been accepted, it packetizes the ac-3 bitstream and sends it to the server. At the server, the received packets fill a buffer. The client is blocked from sending more packets until the server requests for more packets. At the server, the ac-3 bitstream is parsed to extract MDCT coefficients. The extracted MDCT coefficients are projected onto a basis (obtained from offline analysis during training). The likelihood of resulting feature vector is computed under competing Gaussian Mixture Models (GMMs) for each sound class. The sound class GMM with maximum likelihood is declared the winner. After such a classification of every incoming audio frame, one can compute the percentage of certain important audio class that is indicative of “interesting” events for that genre of content. For instance, it turns out that for sports content the important audio class that marks interesting events is a mixture of commentator’s excited speech and audience cheering. The server performs all this analysis from the ac-3 stream in the buffer and whenever the buffer is read completely, it requests for more packets from the client.

Figure 4 shows the timing diagram for the client-server implementation.

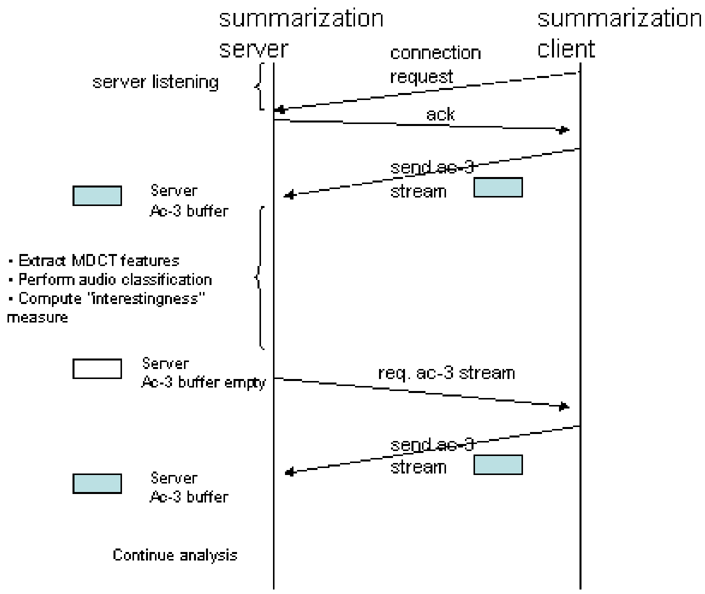

Figure 4: Timing diagram for a client-server implementation of summarization service

5. CONCLUSIONS

In this paper, we proposed two architectures for deploying summarization service over a network. This is an alternative choice to having the analysis algorithms being built into the digital video recorders. The architectures are motivated by the increasing availability of computational resources and usable bandwidth at homes. The digital video recorder offloads the analysis for summarization by streaming the bit-stream to be analyzed to a summarization server. We described an example client-server implementation for summarization service in a home network.