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

A Content-Based Browsing System for a HDD and/or Recordable-DVD Personal Video Recorder

Nakane, K.; Otsuka, I.; Esumi, K.; Divakaran, A.; Murakami,T.

TR2003-89 July 2003

Abstract

The Personal Video Recorder such as Recordable-DVD Recorder and/or Hard Disk Recorder has become popular as a large volume storage device for video/audio content and a browsing function that would quickly provide a desired scene to the user is required as an essential part of such a large capacity system. We propose an intra-program content browsing system using not only a combination of motion based video summarization and topic-related metadata in the incoming video stream but also an audio-assisted video browsing feature that enables completely automatic topic-based browsing.

Publication History:

1. First printing, TR-2003-89, July 2003

A CONTENT-BASED BROWSING SYSTEM FOR A HDD AND/OR RECORDABLE-DVD PERSONAL VIDEO RECORDER

Kazuhiko Nakane [1], Isao Otsuka [1], Kenji Esumi [1], Ajay Divakaran [2] and Tokumichi Murakami [1]

[1] Advanced Technology R&D Center, Mitsubishi Electric Corporation, KYOTO, JAPAN

[2] Mitsubishi Electric Research Laboratories, Murray Hill, NJ

Abstract

The Personal Video Recorder such as Recordable-DVD Recorder and/or Hard Disk Recorder has become popular as a large volume storage device for video/audio content and a browsing function that would quickly provide a desired scene to the user is required as an essential part of such a large capacity system.
We propose an intra-program content browsing system using not only a combination of motion based video summarization and topic-related metadata in the incoming video stream but also an audio-assisted video browsing feature that enables completely automatic topic-based browsing.

1. Introduction

The Personal Video Recorder (PVR) such as Recordable-DVD Recorder and/or Hard Disk (HDD) Recorder has become popular for a storage device of large volume video/audio contents [1].

A browsing function that would quickly provide a desired scene to the user is required as an essential part for such a large capacity system.

In our previous work [2], we proposed an intra-program content browsing system using a combination of motion based video summarization and topic-related metadata in the incoming video stream.

The problem with this approach is that such metadata is often unavailable thus making automatic topic-based browsing impossible. In this paper, we propose an audio-assisted video browsing feature for the recorder that enables completely automatic topic-based browsing as well as extraction of sports highlights using audio features. Furthermore, we describe a preliminary user interface for the proposed browsing feature.

2. System Configuration

A simplified block diagram of the investigated content-based browsing system is shown in Figure 1.

The video stream is encoded by the MPEG Encoder and recorded on the disk through the Buffer controller. The recorded stream is read out from the disc to the MPEG Decoder and the MPEG decoded stream is merged with the input video stream at the Graphics processor for a browsing interface.

The MPEG Encoder is enhanced to implement a function of extracting (1) Color dc Values and (2) Motion Vectors for each block using compressed domain features. We also employ (3) audio features to label each audio segment as one of several classes.

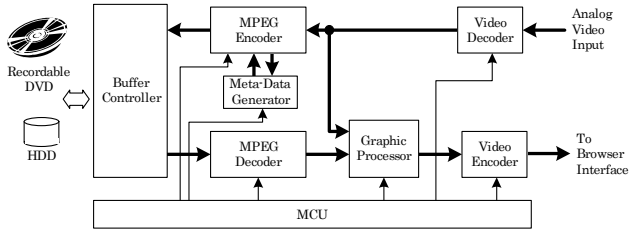

Figure 1. Simplified Block Diagram

3. Audio-assisted Video Browsing Feature

We describe a video summarization technique based on sampling in the "cumulative motion activity space." The intuition is that segments with high motion should require more key-frames for summarization than segments with low motion.

We have devised a scheme that provides a unique and rapid way to first find out the required number of key-frames and then compute them. The visual summary of the entire set of segments is then the concatenation of their key-frames. This method works best when the semantic segment boundaries of the content are known.

We employ the audio features to label each audio segment as one of seven classes, for example, male speech, female speech, applause, cheering, ball hits, speech with music and music, using low-complexity Hidden Markov Models (HMM). We also generate a feature vector for every segment that consists of the state duration histogram corresponding to the HMM for its audio class. The audio segment labels are useful in

sports video to locate the highlights since we look for segments with long sections of applause.

We then use motion vector features to weed out false alarms. With news video, we use the audio feature vectors of the male and female speech segments to identify the principal speakers using simple clustering approach. Our accuracy is moderate but is promising and can be improved. Since in news video, each topic begins with a principal speaker, we can thus easily find the beginning and end of each topic. This allows us to locate the semantic boundaries of the content segments.

We can then summarize each audio segment using our motion-based summarization. We illustrate our approach in Figures 2.

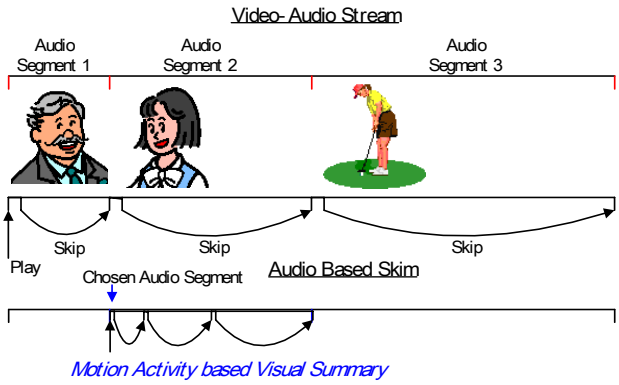

Figure 2. Audio Assisted Video Browsing

In Figure 1, we illustrate the simple incorporation of the proposed browsing approach into the recorder or player. Note that the simple feature extraction allows all the meta-data generation to be done at the encoder side in one pass, i.e. the motion activity, the color dc values and the audio segment labels and feature vectors. The meta-data is then written out in a separate section in the DVD or HDD. The browser then carries out a second pass on the meta-data to identify highlights and principal speakers. Note that the second pass is computationally simpler and thus can be handled by the browser.

4. User Interface Model

The user-interface issue, at the time of applying the content-based browsing system to consumer PVR, is tackled by seamless integration of the browser and player as illustrated in Figure 3 and Figure 4. The interface provides both "Summarization Playback" and "Key-Frame Skip" functions.

"Summarization Playback" is constructed by playing the video sequence starting from the key-frame for a specified duration, and then skipping to the next key-frame until all the key-frames have been displayed.

We also provide "Key-Frame Skip" which enables jumping from a key-frame to the next/previous key-frames or skipping to a desired key-frame so as to give the user another powerful way to browse the content.

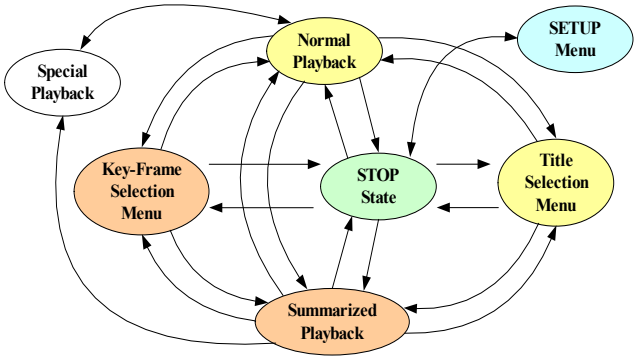

Figure 3. State Transition by User Operation

Figure 4. Image of Key-Frame Selection Menu

5. Conclusion

We have implemented a prototype on personal computers. The simplicity of our approach motivates us to investigate implementation on a set-top box, and in DVD player/recorders. We expect that the both the hardware and firmware would require simple enhancements to realize our approach on such platforms.

Acknowledgment

The authors greatly appreciate the valuable contribution of Dr. Masaharu Ogawa and Mr. Regunathan Radhakrishnan to studying and prototyping the system described in this paper.

[1]K.Nakane, Y.Sato, Y.Kiyose, M.Shimamoto and M.Ogawa, "Development of Combined HDD and Recordable DVD Recorder-Player," ICCE), June 16-20, 2002, Los Angeles.

[2]A.Divakaran and R.Cabasson, "Content Based Browsing System for Personal Video Recorders," ICCE, June 16-20, 2002, Los Angeles.