Fast Scene Change Detection for Personal Video Recorder

Jung-Rim Kim, Sungjoo Suh, and Sanghoon Sull, *Member, IEEE*
Department of Electronics and Computer Engineering, Korea University, Seoul, Korea

***Abstract*—In this paper, we propose a fast scene change detection algorithm suitable for personal video recorders with a low powered CPU using reduced images of intra pictures in the MPEG-2 compressed domain.**

I. INTRODUCTION

Currently available personal video recorders (PVRs) comprise a low-powered CPU, a MPEG-2 decoder and local storage as shown in Fig. 1. A PVR records digital video streams into the local storage. Thus, it will be convenient if PVR users are provided with a variety of advanced features such as content-based retrieval, video browsing, video editing and skip-play all of which can be efficiently implemented by first detecting scene changes. In this paper, we propose a fast scene change detection method suitable for a PVR with a low-powered CPU.

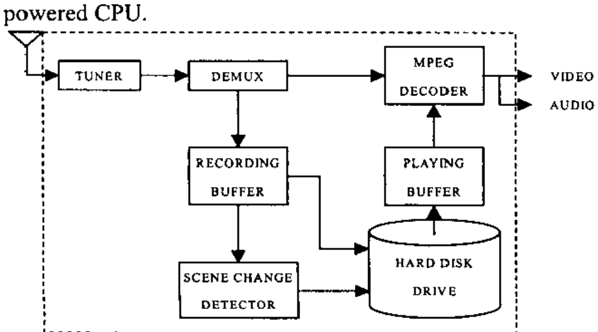


Fig. 1. Basic PVR system.

II. RELATED WORKS

A number of scene change detection methods have been proposed [1-2]. Some of them are based on the correlation of consecutive frames that is not suitable for current low-powered PVRs because the digital broadcast streams generally have high resolution, 720x480 for standard definition television (SDTV) stream and 1920x1080 for high definition television (HDTV) stream. Seong *et al.* [3] recently proposed a scene change detection method based on the number of motion vectors. Although Seong's method requires less computation than previous methods, it is still difficult to be efficiently implemented on low-powered PVRs due to the need for parsing macroblocks in bidirectionally predictive-coded pictures or B-pictures, especially for high-definition video streams that have about six times more macroblocks than standard-definition streams. It also cannot be applied to the high quality video streams which are composed of only intra-coded pictures (I-pictures) and predictive-coded pictures (P-pictures) without B-pictures.

It is observed by the extensive analysis of a number of MPEG-2 video streams that the use of histograms of I-pictures reduced by a factor of four is practically enough for the purpose of scene cut detection. Although scene change can occur in P-picture or B-picture, PVR system may practically start decoding the first I-picture next to the scene change. Thus, in this paper, we present an efficient method based on the observation.

III. PROPOSED ALGORITHM

A. Generation of Reduced Images

The digital broadcast streams are being transmitted in the form of MPEG-2 transport streams of interlaced scanned videos and the PVR system can start displaying the stream from I-picture. Thus, from the practical point of view, we generate the reduced images only from the I-pictures.

In the case of interlaced scanned video, one frame picture consists of two images called top field and bottom field obtained at the different instants of time. MPEG-2 video standard supports the mixture of frame-mode and field-mode macroblocks. A field-mode macroblock has normally larger motion than a frame-mode macroblock. Thus, we construct the partially decoded reduced images in two ways according to the macroblock coding type. For a frame-mode macroblock shown in Fig. 2(a), we use a simple DCT truncation method. For a field-mode macroblock shown in Fig. 2(b), we use only either top field blocks or bottom field blocks. For example, when only top field blocks are used, two upper 8x8 blocks representing top fields in a macroblock are reduced to two 4x2 blocks by reducing the horizontal resolution by four and the vertical resolution by two. The resulting equation taking 4-point and 2-point IDCT in series [4] is as follows:

$$R_{ij} = \frac{1}{8}\sum_{u=0}^{3}\sum_{v=0}^{1} a(u)b(v)y_{uv}\cos\frac{(2i+1)u\pi}{8}\cos\frac{(2j+1)v\pi}{4},$$

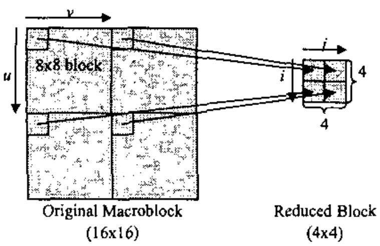
for $i = 0, 1, 2, 3,\ j = 0, 1,$

where $a(u)=\begin{cases}1, & u=0\\ \sqrt{2}, & otherwise\end{cases}$, $b(v)=\begin{cases}1, & v=0\\ \sqrt{2}, & otherwise\end{cases}$

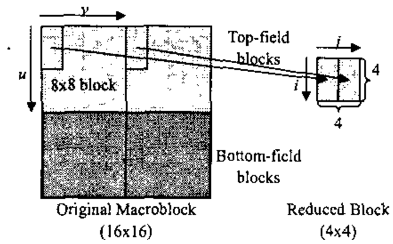
and R_{ij} and y_{uv} represent pixel value in the reduced block and DCT coefficients in original 8x8 DCT block, respectively.

By using only either top field blocks or bottom field blocks, artifacts such as blurring are avoided when there is a rapid motion between top and bottom field. Thus it helps detecting field-overlapped shot boundaries as shown in Fig. 3.

0-7803-7721-4/03 $17.00 © 2003 IEEE



(a) Frame-mode macroblock.



(b) Field-mode macroblock.

Fig. 2. Reduced image generation.

Fig. 3. Field-overlapped shot boundaries.

B. Scene Change Detection

Since current PVR system has very low-powered CPU, it is desirable to extract very simple feature vector from the reduced image for scene change detection. Thus, we generate the histograms, H_n, for both luminance and chrominance from the n-th reduced image. In our experiments, each color channel of YCbCr is uniformly quanitzed to 16 bins and 4 bins, respectively, resulting in a 256 dimensional color feature vector. The L_1 distance of the histogram is defined as follows:

$$d_n = \sum_{k=0}^{255} | H_n(k) - H_{n-1}(k) |,$$

where $H_0(k) = 0$ for $0 \le k \le 255$.

Finally, we get the scene changes S using a threshold T of the distance of histogram:

$$S = \{n \mid d_n \ge T, n = 1,2,3,\ldots,N\}.$$

IV. EXPERIMENTAL RESULTS

To evaluate the performance, we tested our proposed algorithm for a number of MPEG-2 video streams on a system with 1.7 GHz Pentium CPU and 1 GB memory. Table I shows the shot detection accuracy for three video streams (drama, news, and sports) that are all about 10 minutes long and interlaced coded with 720 x 480 and 15 frame GOP. The proposed algorithm is slightly better than Seong's method. We tested our algorithm on a commercial PVR with 150 MIPS PowerPC CPU, 16 MB memory and the embedded Linux. Table II and Table III show the processing time of two methods on a PC system and a commercial PVR system, respectively. The proposed scene change detection algorithm appears two times faster than Seong's method. We also applied both our algorithm and Seong's algorithm to a 4-minute-long 1080i HDTV stream. Our method takes 36 seconds whereas Seong's takes 74 seconds on PC system. Further, Seong's method does not work well for HDTV streams containing the large number of motion vectors and field-mode macroblocks.

TABLE I
SHOT DETECTION ACCURACY

		Drama	News	Sports
Seong's method	Precision	0.877	0.815	0.825
	Recall	0.853	0.821	0.785
Proposed method	Precision	0.960	0.879	0.852
	Recall	0.960	0.829	0.852

TABLE II
PROCESSING TIME ON PC SYSTEM.

	Drama	News	Sports
Seong's method	44 sec	43 sec	43 sec
Proposed method	22 sec	18 sec	18 sec

TABLE III
PROCESSING TIME ON A COMMERCIAL PVR SYSTEM.

	Drama	News	Sports
Seong's method	581 sec	586 sec	596 sec
Proposed method	294 sec	308 sec	309 sec

V. CONCLUSION

In this paper, we have proposed an efficient scene change detection algorithm suitable for a low-powered PVR. The algorithm is based on the histogram difference of the partially decoded intra pictures. The experimental results showed that the proposed algorithm performed slightly better and about two times faster than the previous method.

REFERENCES

[1] A. Nagasaka and Y. Tanaka, "Automatic video indexing and full-motion search for object appearance," in *Proc. IFIP TC2/WG2.6 Second Working Conference on Visual Database Systems*, pp.113-127, 1991.

[2] B. Yeo and B. Liu, "Rapid scene analysis on compressed video," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 6, pp. 533-544, 1995.

[3] Y. Seong, Y. Choi, J. Park, and T. Choi, "A hard disk drive embedded digital satellite receiver with scene change detector for video indexing," *IEEE Transactions on Consumer Electronics*, vol.48, no.3, pp.776-782, 2002.

[4] R. Dugad and N. Ahuja, "A Fast Scheme for Image Size Change in the Compressed Domain," *IEEE Transactions Circuits and Systems for Video Technology*, vol. 11, no. 4, pp. 461-474, Apr. 2001.