A Robust Scene Change Detection Using Mode Distribution in H.264/AVC

Young-Suk Yoon, Won-Young Yoo, and Young-Ho Suh
Contents Protection & Management Research Team
Electronics and Telecommunications Research Institute
138 Gajeongno, Yuseong-gu, Deajeon, 305-350, Korea
{ys.yoon, zero2, syh}@etri.re.kr

***Abstract*—In this paper, we propose a novel scene change detection (SCD) scheme which is available for a sematic video retrieval technique. Using the rate-distortion optimization (RDO) technique used in the H.264 reference software, we have developed an efficient SCD scheme based on the analysis of the mode distribution between intra modes and inter modes. In order to enhance the accuracy of detecting the scene changes, we have also modified the RD function used in RDO technique. Simulation results on several digital videos including abrupt and gradual scene changes show that the proposed scheme provides enhanced performance over previous works.**

***Keywords*-Scene change detection; Video retrieval; Mode distribution; H.264/AVC**

I. INTRODUCTION

Multimedia users would like to search a digital video trying to find out in a lot of related digital videos and wish to be recommended ones similar to a query video. Moreover, contents providers need to protect their digital videos from illegal users in order to avoid an infringement of copyright. However, it is difficult to manage and handle massive amount of digital videos including many frames. Therefore, we need to analyze digital videos into their features.

In general, a video sequence can be divided into spatial and temporal features for the efficient analysis such as browsing, indexing, retrieving, monitoring, editing, and authoring digital video. First, the spatial feature depicts edge, texture information, and spatial complexity of a frame. Next, the temporal feature represents the time continuity and discontinuity for scenes, the motion of objects, and optical flows. Especially, a scene change expresses the gap between a scene and the next another for a digital video. A set of scene change is the sematic and reliable information which is used to differentiate videos from each other.

Fig. 1 illustrates frames of a digital video based on a time domain. Video sequences consist of a lot of consecutive frames. A scene is a set of frames connected according to a semantic context of a digital video. Herein, $Scene_{n-1}$ and $Scene_n$ have semantically different frame configurations and contexts, and a scene change exists at a boundary between two scenes.

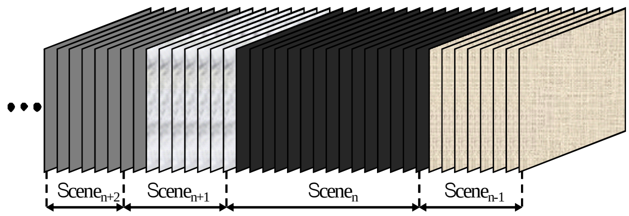


Figure 1. Hierarchical structure of a video sequence

Many researchers have proposed various methods for scene change detection such as pixel-based, histogram-based, edge-based, statistics-based, compression-based, and hybrid methods. Pixel-based methods used the difference of pixel values between successive frames [1]. Gray or color histogram-based schemes compared histograms for neighboring frames, respectively [2]. Edge-based methods employed either object segmentation or edge detection scheme and estimated the degree of change compared with outlines of consecutive frames [3]. Statistics-based methods employed many statistical inferences used in signal processing [4]. Compression-based methods utilized a concept which describes more information is needed to encode a frame when there is a scene change [5]. Hybrid methods apply more than two methods to scene change detection in order to obtain better detecting performance [6].

However, it is not easy to correctly detect a scene change which is necessary for a semantic-oriented video retrieval. In this paper, we present H.264/AVC based mode type [7] classification method for reliable detection of various scene changes. Analyzing the mode distribution between inter and intra modes, we measure temporal and spatial correlation of each frame and utilize the correlation ratio to determine a scene change. Furthermore, we have modified the RD function by using mean removed sum of squared difference (MRSSD) in order to achieve robust scene change detection for illumination change. For the scene change detection between two similar background shots or simple background shots, we have also proposed a selective mode counting (SMC) technique. Moreover, we do not consider the mode information of macroblocks which regions are homogenous or are located in the boundary area.

II. MODE DECISION IN H.264/AVC

The latest video coding standard, H.264/AVC, uses variable block sizes ranging from 4×4 to 16×16 in interframe coding. To achieve the highest coding efficiency, H.264/AVC uses rate-distortion optimization (RDO) technique which maximizes coding quality and minimizes resulting data bits. The RDO mode decision method finds the optimal prediction mode in terms of rate distortion. This method computes rate-distortion (RD) cost based on the actual rate and distortion after successive processes, transform, quantization, entropy coding, and reconstruction. The RD cost is defined [8] as

$$J(s,c,M \mid QP,\lambda_{MODE}) = SSD(s,c,M \mid QP) + \lambda_{MODE} \cdot R(s,c,M \mid QP) \quad (1)$$

where s and c are the source video signal and the reconstructed video signal, respectively. QP is the quantization parameter and λ_{MODE} is the Lagrange multiplier. $SSD(s,c,M \mid QP)$ is the sum of the squared differences between s and c. M indicates a MB mode. In Eq. (1), $R(s,c,M \mid QP)$ is the number of bits associated with the given M and QP. H.264/AVC encoder calculates the RD cost of every possible mode and chooses the best mode having the minimum RD cost.

H.264/AVC adopts the highest number of modes than any other video coding standards. For a P frame, a macroblock can be coded in the middle of the possible modes {SKIP, 16×16, 16×8, 8×16, 8×8, 8×4, 4×8, 4×4, I4×4, I16×16}. We, thus, classify those mode set into two categories, such as inter mode and intra mode as follows:

INTER MODE {SKIP, 16×16, 16×8, 8×16, P8×8}
INTRA MODE {I 16×16, I 4×4}

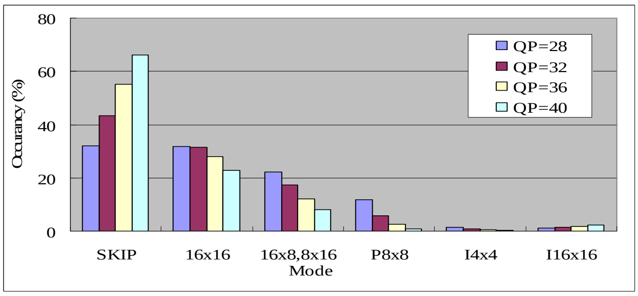
III. PROPOSED METHOD

In the conventional scene change detection algorithms including [5], the sequence of sum of absolute difference (SAD) values between frames have been computed and used to detect scene changes. In addition, they have also considered statistical properties, such as mean value and standard deviation used to define a continuously updating automated threshold. In general, they make a decision for scene change when a high SAD value is observed between frames. However, as shown in Fig. 2, we can observe high SAD values during rapid movement, abrupt illumination changes and transition effects such as zoom in/out, fade in/out, dissolve etc. Moreover, we can have scene changes with very different value levels. A scene break, where both scenes have similar background, does not give a peak as high as if they had different ones. Consequently, a proper decision function is needed, which can take into account characteristics of the scene without a previous input.

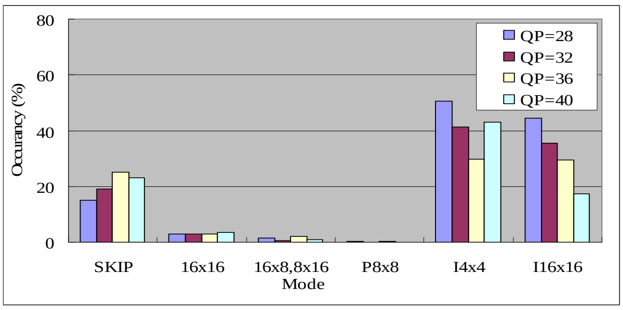
Figure 2. Various kinds of scenes from the test sequences

For a new decision function, we utilize the RDO technique in H.264/AVC because the distribution of the best mode obtained from RDO effectively represents the relationship between temporal correlation and spatial correlation in each frame. Since temporal correlation is even higher than spatial correlation for a frame which does not belong to scene change. *INTRA MODE* rarely occurs (about 2%) in the inter frame as shown in Fig. 3. However, *INTRA MODE* frequently occurs where scenes change. These characteristics have already been investigated and verified in [7] and our previous research work [8].

Fig. 3 shows the best mode distribution according to the existence of scene changes. Therefore, using mode distribution between *INTRA MODE* and *INTER MODE*, we can efficiently find the exact time positions where the real scene changes occur.



(a) Mode distribution in non-scene change frame



(b) Mode distribution in scene change frame

Figure 3. The mode distribution

A. Propocessing

First of all, the proposed system with video decoders using FFmpeg software [9] decodes input video data into frame sequences as shown in Fig. 4. It then normalizes temporally decoded sequences and spatially resized frames to detect scene change in the same circumstances (Frame size: 320x240, Frame rate: 10fps). Temporal normalization reduces processing time and diminishes effect of continuous scene change. Furthermore, spatial normalization decreases computational complexity to detect scene change.

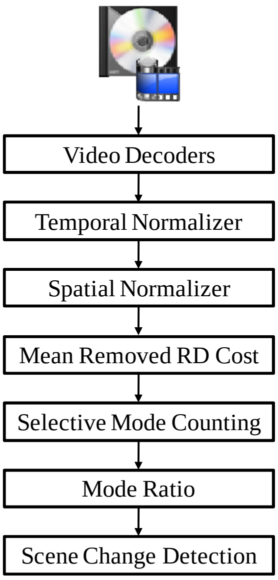


Figure 4. Block Diagram of Proposed Method

B. Mean Removed RD Cost

The severe illumination changes seem to be scene-cuts and increase the number of false hits. In order to design robust scene change detection scheme for illumination change, we have modified the RD function by using mean removed sum of squared difference as follows

$$MRSSD = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\{(x_{ij} - m_x) - (y_{ij} - m_y)\}^2 \qquad (2)$$

where x_{ij} and y_{ij} represent the pixel intensity of original block and motion compensated block, respectively. In Eq. (2), m_x and m_y are the average pixel intensity in each block.

$$m_x = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} x_{ij}, \quad m_y = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N} y_{ij} \qquad (3)$$

Therefore, we have changed the RD cost function using Eq. (1) and (2) as

$$J(s,c,M \mid QP,\lambda_{MODE}) = MRSSD(s,c,M \mid QP) + \lambda_{MODE} \cdot R(s,c,M \mid QP) \qquad . (4)$$

C. Selective Mode Counting (SMC)

For the scene change detection between two similar background shots or simple background shots as shown in Fig. 2, we use a selective mode counting (SMC) technique. It does not consider the mode information of macroblocks; the regions of which are homogenous or are located in the boundary area. The homogeneity of a macroblock is checked by RD cost as follows

$$J(s,c,M \mid QP,\lambda_{MODE}) < \overline{J_P(s,c,SKIP \mid QP,\lambda_{MODE})} \qquad (5)$$

where $\overline{J_P(s,c,SKIP \mid QP,\lambda_{MODE})}$ represents the average RD cost for SKIP mode in a previous frame. The boundary area in each macroblock is simply regarded as lines located in the most right, left, top, and bottom of each frame.

D. Mode Ratio (MR) and Scene Change Detection

Finally, we can obtain the mode ratio (MR) from Eq. (5) and determine that a scene change occurs if MR is larger than the given threshold (*TH*). For simplicity, we have fixed the *TH* into 60 in our experiments.

$$MR = \frac{Count\_INTRA}{Count\_INTER + Count\_INTRA} \times 100 > TH \qquad (6)$$

In Eq. (6), *Count_INTRA* and *Count_INTER* are the number of valid intra modes and inter modes, respectively.

IV. EXPERIMENTAL RESULTS

In this section, we validate the proposed scheme for detecting various scene changes. Simulations were carried out using H.264/AVC reference software JM 12.4 [10] and coding parameters used are shown in Table 1. However, the proposed system does not encode frame sequences into bit streams for H.264/AVC, but utilizes only the information of inter and intra modes.

TABLE 1 SIMULATION CONDITIONS

Reference Software	JM 12.4 [10]
Profile	Baseline
RDO Mode	Fast High Complexity Mode
GOP Structure	I P P P · · ·
Reference Frames	2
Search Range	±16
FME	UMHexagonS

We used video sequences of music video and commercial advertisement contents as shown in Fig. 2. They were chosen because they have scenes with intense motion, change of light conditions, high complexity and different types of scene changes.

Its resolution is 320x240 and its length is 2,500 frames. The number of true scene changes in this sequence was 70.

 ISBN: 978-1-61208-240-0

We obtained them ‘manually’ by watching the video and counting them.

The performance of the proposed method is evaluated by comparing with other methods and the ground truth. For this reason, the “recall” and “precision” ratios are defined as follows

$$R = \frac{N_c}{N_c + N_M} \times 100\ (\%) \tag{7}$$

$$P = \frac{N_c}{N_c + N_F} \times 100\ (\%) \tag{8}$$

where N_C, N_F, and N_M are the number of correct detections, the number of false ones, and the number of missed ones, respectively.

Table 2 shows the results of the various scene change detection algorithm. For the first two cases, after having the SAD values for the whole sequence, the fixed threshold and Dynamic threshold in [5] were chosen optimally to minimize the number of missed and false detections. For the last two cases, RD function in H.264/AVC and our proposed schemes are used, respectively.

TABLE 2 PERFORMANCE OF PROPOSED METHOD

	N_C	N_F	N_M	R(%)	P(%)
Fixed Threshold	56	29	14	65.88	80.00
Adaptive Threshold	60	17	10	77.92	85.71
H.264/AVC RD	60	9	2	96.77	86.95
Proposed Method	64	5	1	98.46	92.75

In Fig. 5, we also represent the distribution of SAD and MR for each frame from the experimental results. We can verify that the mode type classification technique using RD function is more appropriate to detect various scene changes than the conventional approach using SAD function. We also confirmed that our proposed method using MRSSD, SMC and MR enhances the recall and precision ratio about 2% and 6% compared with RDO technique in H.264/AVC, respectively.

V. CONCLUSION

In this paper, we presented a reliable detection method for various abrupt and continuous scene changes through an analysis of the mode distribution between intra modes and inter modes in each frame. In order to enhance scene change detection ratio, we have adopted mean removed sum of squared difference (MRSSD), selective mode counting (SMC) and mode ratio (MR) schemes. Based on these schemes, the proposed scene change detection technique works better than others in detecting dissolves with low variance frames, and decreases false hits induced by illumination change.

ACKNOWLEDGEMENT

This research project was supported by the Government Fund from Korea Copyright Commission. [2012-cloud-9500 : Development of content-based usage control technology for clean cloud]

REFERENCES

[1] H. J. Zhang, A. Kankanhalli, and S. W. Smoliar, “Automatic Partitioning of Full-Motion Video,” *Multimedia Systems*, vol.1, no.1, pp. 10-28, 1993.

[2] C. F. Lam and M. C. Lee, “Video Segmentation using Color Difference Histogram,” *Lecture Note in Computer Science 1464*, New York: Springer-Verlag, pp. 159-174, 1998.

[3] W. J. Heng and K. N. Ngan, “High Accuracy Flashlight Scene Determination for Shot Boundary Detection,” *Signal Processing: Image Communication*, vol.18, no.3, pp. 203-219, Mar. 2003.

[4] I. K. Sethi and N. Patel, “A Statistical Approach to Scene Change Detection,” *Proceedings of SPIE*, vol. 2420, pp. 329-338, Feb. 1995.

[5] A. Dimou, O. Nemethova, and M. Rupp, “Scene Change Detection for H.264 using Dynamic Threshold technique,” *Proceedings. of 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Service*, Jun. 2005.

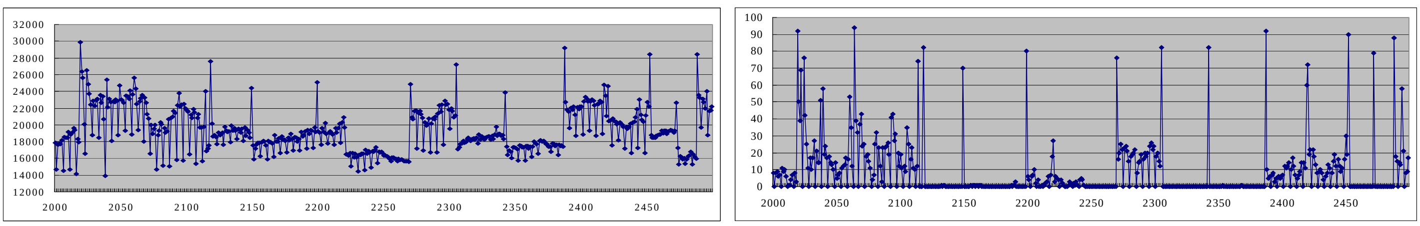
[6] C. L. Huang and B. Y. Liao, “A Robust Scene-Change Detection Method for Video Segmentation,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol.11, no.12, pp. 1281-1288, Dec. 2001.

[7] D. S. Turaga and T. Chen, “Estimation and Mode Decision for Spatially Correlated Motion Sequences,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol.11, no.10, pp. 1098-1107, Oct. 2001.

[8] S. H. Kim and Y. S. Ho "A Fast Mode Decision Algorithm for H.264 using Statistics of the Rate Distortion Cost," Electronics Letters, vol. 44, no. 14, pp. 849-851, 2008.

[9] FFmpeg Revision 13712, available online at: http://ffmpeg.mplayerhq.hu/.

[10] JVT reference software version 12.4, available online at: http://iphome.hhi.de/suehring/tml/download/old_jm/.



(a) Distribution of the sum of absolute difference between frames

(b) Distribution of the mode ratio in each frame

Figure 5. Distribution of SAD and MR