Scene Change Detection on H.264/AVC Compressed Video Using Intra Mode Distribution Histogram Based on Intra Prediction Mode

MINYOUNG EOM and YOONSIK CHOE
Department of Electrical and Electronics Engineering, Yonsei University
134 Sinchon-dong Seodeamoon-goo, Seoul, 120-749 KOREA

Abstract: - This paper presents a novel scene change detection method using complexity and edge direction based on intra prediction mode in H.264/AVC compressed video. When scene change occurs, there is less temporal correlation between consecutive frames. Therefore, most of macro-blocks are encoded in intra mode. Using this property, the proposed method calculates the percentage of intra mode blocks in each predictive frame in order to get candidates of scene change frames. To detect accurate scene change frames, we generate the Intra Mode Distribution Histogram (IMDH) representing edge directionality and image roughness. By similarity matching of consecutive two IMDHs, we decide scene change frames. The distance between two IMDHs is measured using χ^2-statistic. The experimental results show that our method is efficient and robust.

Key-Words: Scene Change Detection, H.264/AVC, Intra Mode Distribution Histogram.

1 Introduction

Due to rapid development of data communication techniques and network environment, the amount of multimedia data has been increased. The requirements for efficiently accessing mass amounts of multimedia data are becoming more and more important. Video scene change detection is a fundamental operation used in many multimedia applications such as digital libraries and video on demand (VOD), and it must be performed prior to all other processes. However, because most of all videos are stored and transmitted as compressed form such as MPEG-1/2/4 or H.264, decoding procedure is essentially needed. It is time consuming as well as computationally expensive and makes real time process impossible. In order to solve these problems, there are many researches for the scene change detection in compress domain. For example, in [1], J. Kim et al. proposed DC coefficients based approach, Dawood A.M et al. utilized the motion vector for the scene change detection [2], Li H. et al. used the bit rate of intra frame and S.W. Lee et al. used edge orientation and strength extracted by AC coefficients analysis [3][4]. However, these methods mostly focus on the MPEG-1/2 video coding standard.

H.264/AVC is the newest high compression digital video codec standard having several different coding schemes such as intra prediction, 4×4 integer transform, various sizes of MB (Macro Block) and multiple references. These different coding schemes make to directly use of several scene change detection schemes in MPEG1/2 impossible. In other words, H.264/AVC adopts intra prediction for intra frame. So the approach using DCT coefficients of intra frame is difficult to be adapted to H.264 video. Motion vector based approach is also same due to various sizes of MB and multiple references in H.264.

The latest, the research of the scene change detection for H.264/AVC video has been studied. In [5], S. Kim et al. proposed the scheme in H.264/AVC compression domain which uses prediction mode variations between consecutive two intra frames. However, it has a drawback in detecting accurate scene change frames because it considers only intra frames.

In this paper, we propose an accurate scene change detection method that uses the local distribution of edge directionality and image roughness based on intra prediction mode. To do this, we define a new feature histogram, what we called Intra Mode Distribution Histogram (IMDH), which represents directionality and roughness in local area of a frame image. After we decide a scene change candidate frame using the number of intra coded blocks, we generate IMDH in candidate frames. Finally, we can obtain accurate scene change frames by similarity matching using χ^2-statistic which is a formal statistical method for determining whether two distributions differ

This paper is organized as follows. Section 2 presents the former method. In section 3 the definition of IMDH is defined and its semantics is described. Intra mode distribution histogram extraction method and proposed scene change detection scheme is introduced in section 4. In section 5 we compare the previous method with our proposal method via experiment. Then we conclude this paper.

2 Previous Scene Change Detection

Intra prediction is one of the unique features newly adapted to H.264/AVC. The prediction mode for intra frame is classified into *16×16* and *4×4*. That is, while the macro blocks with *16×16* prediction mode have monotonous gray levels in the block, those with *4×4* prediction mode show large dynamic range in the gray levels. Therefore, there will be a large variation in the prediction mode distribution between consecutive two intra frames with a scene change [5].

Let $Mode_i^j$ indicate the intra prediction mode for *j*th MB in the *i*th frame:

$$Mode_i^j = \begin{cases} 0 & \text{for } 16\times 16 \text{ prediction mode} \\ 1 & \text{for } 4\times 4 \text{ prediction mode} \end{cases} \qquad (1)$$

Then, for the similarity measure of two consecutive *i*th and *(i+1)*th intra frames, they define the normalized feature Δf_i^k . It indicates the normalized prediction mode variations of each sub-block (there are 20 non-overlapping sub-blocks in a frame) between *i*th and *(i+1)*th intra frame as follows:

$$\Delta f_i^k = \frac{1}{|S^k|}\left| \sum_{j\in S^k} Mode_i^j - \sum_{j\in S^k} Mode_{i+1}^j \right| \qquad (2)$$

where S^k is a set of MBs which are included within the *k*th sub-block of an intra frame, $|S^k|$ is the cardinality of S^k and k is the number of sub-block within a frame.

The decision function is defined as follows:

$$d_{i,i+1} = \frac{1}{N_{sub}} \sum_{\forall k} q_i^k \ , \ q_i^k = \begin{cases} 1, & \Delta f_i^k \ge \lambda \\ 0, & \Delta f_i^k < \lambda \end{cases} \qquad (3)$$

where N_{sub} represent the number of sub-blocks and λ is the predefined threshold which are applied to each sub-block. That is, if $d_{i,i+1}$ is larger than the predefined value T', then there is a scene change.

3 Definition and Semantics of IMDH

The IMDH basically represents the local distribution of 4 types of edge directions and 2 types of roughness based on intra prediction mode in each local area called a sub-image. Specifically, dividing the image space into *4×4* sub-images as shown in Fig. 1, the local distribution for each sub-image can be represented by a histogram. To generate the histogram, intra prediction directions in the sub-images are categorized into four types; vertical, horizontal, 45° diagonal and 135° diagonal and kinds of intra prediction mode are categorized into

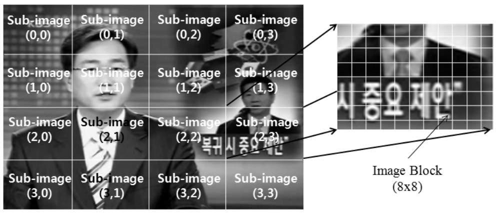

Fig.1. Definition of sub-image and image-block

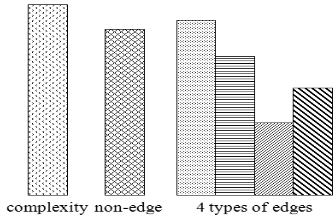

Fig. 2. Six types of IMDH bin in each sub-image

Table 1. Semantics of the histogram bins of the IMDH

IMDH bin	Semantics
h[0]	Relative frequency of intra 4×4 mode in sub-image at (0,0)
h[1]	Relative frequency of DC mode in sub-image at (0,0)
h[2]	Relative frequency of vertical edges in sub-image at (0,0)
h[3]	Relative frequency of horizontal edges in sub-image at (0,0)
h[4]	Relative frequency of 45° edges in sub-image at (0,0)
h[5]	Relative frequency of 135° edges in sub-image at (0,0)
⋮	⋮
h[90]	Relative frequency of intra 4×4 mode in sub-image at (3,3)
h[91]	Relative frequency of DC mode in sub-image at (3,3)
h[92]	Relative frequency of vertical edges in sub-image at (3,3)
h[93]	Relative frequency of horizontal edges in sub-image at (3,3)
h[94]	Relative frequency of 45° edges in sub-image at (3,3)
h[95]	Relative frequency of 135° edges in sub-image at (3,3)

two types. Thus, the histogram for each sub-image represents the relative frequency of occurrence of the four types of directions and two types of roughness in the corresponding sub-image. As a result, as shown in Fig. 2, each local histogram contains 6 bins. Since there are 16 sub-images, a total of 6×16=96 histogram bins are required. Table 1 summarizes the semantics of the 96-bin IMDH. The basic counting unit for generation of the histogram is *8×8* image-block.

4 The Proposed Scene Change Detection Based on IMDH

Due to consideration of only intra frame in the previous method, it has a problem which is difficult about detection of correct position of scene change. We propose the more accurate scene change detection method with considering not only intra frame but also inter frame. In this paper, we get candidates of scene change using a percentage of MBs which are encoded as Intra mode. Then, we generate IMDH of each candidate frames with intra prediction modes. Finally, we perform the scene change detection with χ^2-statistic of two IMDHs

4.1. The scene change candidate detection

When a scene change occurs, there is less temporal correlation between the frames and the majority of MBs will be coded in intra mode. We can exploit these properties for scene change candidate detection. We compute the percentage of intra blocks in each frame and compare against a threshold value in order to decide whether the frame has possibility of scene change or not. The following Table 2 shows the average percentage of intra MBs in each frame and the experiment video consists of 9776 frames and has 101 scene changes.

Table 2. Average percentage of selected mode

	Intra mode (%)	Inter mode(%)
scene change frame	88.68	11.32
otherwise	3.74	96.26

4.2. Extraction of IMDH

To generate IMDH of candidate frames, we define the edge direction and complexity of an *8×8* image-block according to intra prediction mode. Intra prediction mode in H.264/AVC is classified into 4 types of *16×16* prediction mode and 9 types of *4×4* prediction mode as Fig. 3 and 4 [6]. Since the mode decision is performed by minimization of rate-distortion cost function, each intra mode reflects edge direction of each block. Table 3 denotes the relationship between the intra mode and edge direction. Let's consider the determined intra mode from a different standpoint. If MB belongs to homogeneous region, *16×16* prediction mode will be selected. That is, if the MB is coded by *4×4* prediction mode, generally we can say that the MB includes complex region such as edge, texture, etc. It means that the number of *4×4* mode in each sub-image reflects complexity in the sub-image. So the first IMDH bin in each sub-image denotes the normalized value of the number of *4×4* modes. The normalization is performed by dividing by four times of the total number of intra coded MBs in the sub-image. And the second bin denotes the relative frequency of DC mode. Because DC mode is decided in most of homogeneous region, the roughness of each sub-image can be approximately represented by the first two bins.

When the MB is decided as *16×16* prediction mode, we assign the edge direction value of *16×16* prediction mode to each 4×4 block for efficient calculation of the

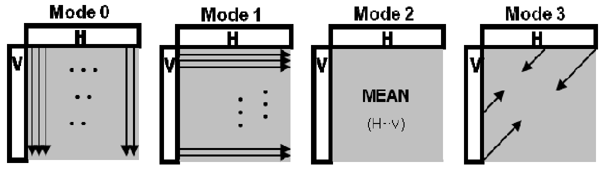

Fig. 3. 16×16 intra prediction mode

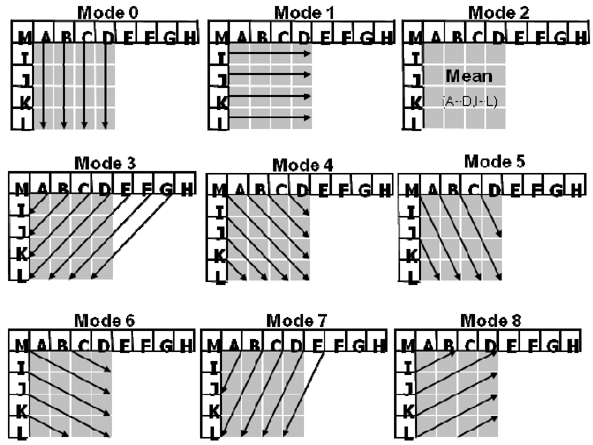

Fig. 4. 4×4 intra prediction mode

Table 3. Edge direction of intra prediction mode

Mode	0	1	2	3
Direction (degree)	90	0 or 180	Non-direction	45
4	5	6	7	8
135	116	153	63	26

edge feature in the image-block. It means that each *4×4* block has a certain edge direction feature. We compute an average edge direction $e_{8\times8}$ of the image-block with the above edge direction value of each prediction mode. That is, $e_{8\times8}$ is obtained by the average of four edge direction values of *4×4* blocks. When calculating $e_{8\times8}$, a *4×4* block of mode 1 is decided that it has 180° directional edge if the dominant edge direction value of the rest of *4×4* blocks is an obtuse angle such as mode 4, 5, and 6. In another case, the edge direction value of mode 1 is decided as 0° if the dominant edge direction value of the rest of the *4×4* blocks is an acute angle such as mode 3, 7, and 8. If every *4×4* block in the image-block is mode 2, $e_{8\times8}$ is decided as non-edge. Otherwise, we don't consider *4×4* block of mode 2. Therefore $e_{8\times8}$ is defined as follows

$$e_{8\times8} = \begin{cases} \dfrac{1}{N_{edge}} \sum_{i=1}^{4} e_{4\times4}^{i} & , \quad e_{4\times4}^{i} \neq non\_edge \\ non\_edge & , \quad e_{4\times4}^{i} = non\_edge \ \forall i \end{cases} \qquad (4)$$

where $e^{i}_{4\times4}$ is the edge direction value of *i*th *4×4* block in the image-block, N_{edge} is the number of *4×4* blocks whose edge direction value is not non-edge.

We define the edge region of the image-block by $e_{8\times8}$. The block is classified into 4 types of the edge region as shown in Fig. 5. In case of the vertical dominant edge, $e_{8\times8}$ belongs to the region 1. If $e_{8\times8}$ belongs to the region 2, the image-block has the horizontal dominant edge. And if $e_{8\times8}$ belongs to the region 3 or 4, then the image-block has 45° orientation or 135° orientation. Therefore, each histogram bin value is counted by $e_{8\times8}$. For normalization, the number of each bin is divided by the total number of image-blocks in the sub-image.

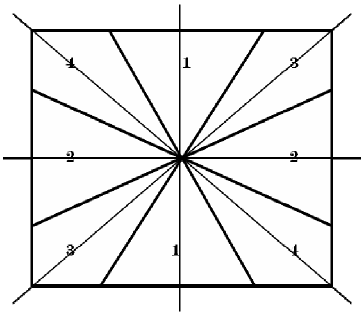

Fig. 5. Edge region classification

4.3. Scene change detection

For the similarity matching to detect scene changes, we calculate the distance $D(H_n, H_{n+1})$ of IMDH for two consecutive frames using the following χ^2-statistic measure:

$$D(H_n, H_{n+1}) = \sum_{i=0}^{95} \frac{(h_n[i] - h_{n+1}[i])^2}{h_n[i] - h_{n+1}[i]} \tag{5}$$

where H_n and H_{n+1} denote the IMDH of two consecutive frames and their *i*th bin values are $h_n[i]$ and $h_{n+1}[i]$, respectively. Finally, if $D(H_n, H_{n+1})$ is larger than the empirical threshold value T, a scene change occurs between frame *n* and *n+1*.

5 Experimental Results

As a criterion for performance evaluation, we used Recall and Precision. Recall and Precision are defined as follows

$$Recall = \frac{N_c}{N_c + N_m} \times 100$$
$$Precision = \frac{N_c}{N_c + N_f} \times 100 \tag{6}$$

where N_c, N_m and N_f are the number of correct, miss, and false detection, respectively. The above equation (6) shows that N_m and N_f have the smaller value, Recall and Precision are the more closely 100% and it means good performance.

Fig. 6. Examples of H.264/AVC compressed video

The test H.264 video set is composed of 15 sequences which are categorized in movie, interview and documentary. Their playtimes are from 30 sec to 90 sec and the size is *352×240*. The Images in Fig. 6 are examples of H.264/AVC compressed video. To compare with the previous method, the experiment is executed with H.264 videos which have one intra frame per 30 frames. The experimental results demonstrate the efficiency and accuracy of the proposed method. Table 4 shows the performance via comparison of the previous method and the proposed method. For some videos, the precision of our method is slightly inferior, but, entirely, the Recall and Precision of the proposed method is greatly superior to the previous method with respect to average Recall/Precision value.

Table 4. Performance comparison between previous and proposed method (previous/proposed)

video (#)	Scene (#)	N_c	N_f	N_m	Precision (%)	Recall (%)
1	11	7/11	1/0	4/0	88/**100**	64/**100**
2	21	15/21	8/0	6/0	64/**100**	71/**100**
3	14	6/14	7/3	8/0	46/**82**	43/**100**
4	37	24/36	8/3	13/1	75/**82**	65/**97**
5	37	29/36	0/3	8/1	100/92	78/**97**
6	12	8/12	0/0	4/0	100/100	67/**100**
7	21	11/21	9/0	10/0	55/**100**	52/**100**
8	27	23/24	6/2	4/3	79/**92**	85/**89**
9	24	16/22	2/0	8/2	89/**100**	67/**92**
10	65	46/62	4/9	19/3	92/87	71/**95**
11	13	11/13	5/1	2/0	68/**93**	85/**100**
12	13	7/10	6/0	6/3	54/**100**	54/**77**
13	11	9/10	9/0	2/1	50/**100**	82/**91**
14	7	6/7	3/0	1/0	67/**100**	86/**100**
15	15	10/13	4/1	5/2	71/**93**	67/**87**
average					73/**95**	69/**95**

5 Conclusion

In this paper, we proposed the scene change detection in H.264/AVC compression domain using intra prediction mode which is a unique feature in H.264/AVC. Firstly, we detected scene change candidates using the average percentage of intra block in a frame. Secondly, with analyzing the intra prediction mode, the intra mode distribution histogram was generated. Finally, we detected the scene changes by the histogram similarity matching using χ^2-statistic. The proposed method can reduce processing time and computations for decoding procedure because it observes only each macro block mode, we don't need a full decoding procedure. Moreover the performance is improved by 26%/22% in terms of average Recall/Precision.

References:

[1]Jungrim Kim, Sungjoo Suh, and Sanghoon Sull, "Fast Scene Change Detection for Personal Video Recoder," *IEEE Trans. on Consumer Electronics*, vol. 49, 2003, pp. 683-688.

[2]Dawood A.M., and Ghanbari M., "Scene Cut Detection from MPEG Video Stream Coded without B Picture," *IEEE International Conference on ICASSP*, vol. 3, 2001, pp. 1645-1648.

[3]Li H., Liu G., Zhang Z., and Li Y., "Adaptive Scene Detection Algorithm for VBR Video Stream," *IEEE Trans. on Multimedia*, vol. 6, 2004, pp. 624-633.

[4]S.W. Lee, Y.M. Kim and S.W. Choi, "Fast Scene Change Detection using Direct Feature Extraction from MPEG compressed videos," *IEEE Trans. on Multimedia*, Vol. 2, No. 4, 2000, pp.240-254.

[5]S.M. Kim, J.W. Byun, and C.S. Won, "A Scene Change Detection in H.264/AVC," *LNCS 3786*, 2005, pp. 1072-1082.

[6]Iain E. G. Richardson, "*H.264 and MPEG-4 Video Compression,*" John Wiley &Sons 2003