2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application

# A Novel Scene Change Detection Algorithm for H.264/AVC Bitstreams

Feng Jie, Huang Aiai, Chen Yaowu
*Institute of Advanced Digital Technology and Instrumentation, Zhejiang University, Hangzhou, 310027, P.R China*
*Email: fengjie_zju@yahoo.cn, aiaizju@163.com, cyw@mail.bme.zju.edu.cn*

## Abstract

*In this paper, we present a novel scene change detection algorithm in the H.264/AVC compressed domain directly. Three major statistical features are used: (i) bit allocation for intra mode macroblocks (BIM), (ii) bit allocation for inter mode macroblocks (BPM), and (iii) the number of skip mode macroblocks in a frame. These features can be easily extracted from the H.264/AVC bitstreams. Besides the percent of skip macroblocks in a frame, an adaptive threshold based on the ratio of BIM to BPM is used to determine the abrupt and the gradual scene changes respectively. Experimental results indicate that the proposed algorithm achieves the good performance with a low computational complexity.*

## 1. Introduction

With the rapid growth of video database, video analysis, indexing, browsing, summarization and retrieval systems become more and more important. The video scene change detection is the first step of these applications. There are two basic types of scene changes. One is the abrupt change where two adjacent frames are completely uncorrelated, and the other is the gradual change where the difference between frames corresponding to two scenes is substantially reduced. There are several gradual change types such as dissolve, fade and wipe. The gradual change is very difficult to be detected since it is usually confused with frames which have large motion, so it will raise a lot of false alarms.

Current works for the scene change detection are mainly using two approaches, which are based on the spatial domain and the compressed domain respectively. In the spatial domain, the histogram comparison is usually used. The main drawback of spatial domain approaches is time consuming because of the uncompressing process and the complexity of the algorithm. In the compressed domain, a lot of features can be extracted from bitstreams without fully decoding it. Many works have been done in MPEG compressed domain. They have used different features such as DC images, DCT coefficients, MB coding modes and motion vectors (MV), and shown the good performance [1].

A few scene change detection techniques in H.264/AVC compressed domain have been proposed in the literature. Kim *et al*. have proposed a prediction mode based scene change detection algorithm in H.264/AVC compressed domain [3]. It only detected the change on I frames and took no account of P and B frames. The weighted prediction tool in H.264/AVC has been used to detect fade in and out by Damghanian *et al*. [4]. Zeng *et al*. used a weighted city block distance of intra mode histograms between two adjacent I frames and the ratio of intra macroblocks for P and B frames [5]. Bruyne *et al*. used two particular features in H.264/AVC, which are the macroblock type and the corresponding display number of the reference frames [6]. They have compared different clips with various GOP lengths and showed very good performance on abrupt change detections. Schoffmann *et al*. used L1-Distance based partition histograms (PHD) to detect abrupt scene changes and the ratio of intra-coded macroblocks (IBR) on a per-frame basis to detect gradual scene changes [7]. It takes only about 10% of the entire decoding time and shows very good performance. But it has two major drawbacks; one is that the threshold was not adaptive and could not adjust to different video clips and the other is that the most gradual changes were detected only by the ratio of intra-coded macroblocks. In fact, all of the above algorithms are mainly based on the fixed threshold to detect the scene changes. However, a fixed threshold algorithm cannot have good performances for all video sequences due to the diversity of the video characteristics. Dimou *et al*. defined an automated, dynamic threshold model [8]. But the Sum of Absolute Differences (SAD) which used in their method could not be extracted in H.264/AVC bitstresams.

As shown in figure 1, our scene change detection algorithm consists of two main processes. Abrupt scene changes are detected firstly, and gradual scene change will be detected later according to the previous result in a four-step process. Comparing with previous algorithms, we achieve better precision and recall values of both abrupt and gradual scene change detection for H.264/AVC bitstreams.

This paper is organized as follows. In Section II, the proposed scene change detection algorithm in H.264/AVC compressed domain is outlined. Section III introduces our algorithms based on the BIM, the BPM and the number of skip mode macroblocks in detail. The experimental results are shown in Section IV. Section V contains a summary and conclusions.

## 2. Scene change detection in H.264/AVC

H.264/AVC is the newest video coding standard of the ITU-T Video Coding Experts Group and the ISO/IEC Moving Picture Experts Group. It offers many new coding tools including intra prediction, variable block size motion compensation, multiple reference picture

978-0-7695-3490-9/08 $25.00 © 2008 IEEE
DOI 10.1109/PACIIA.2008.246

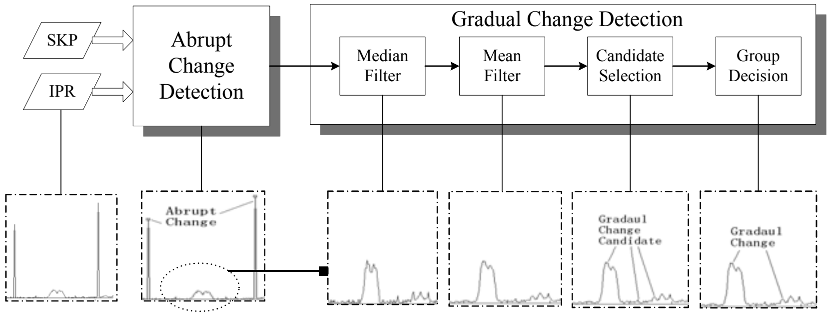
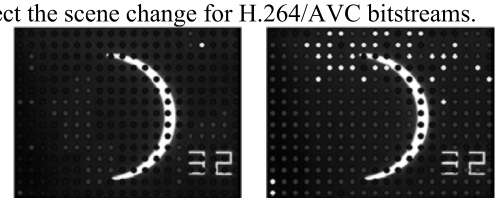



**Figure 1 Overview of the proposed scene change detection algorithm**

motion compensation, weighted prediction, and in-the-loop deblocking filtering which are not used in the previous video coding standards, such as MPEG-2 and MPEG-4 [2]. Due to its superior compression performance, more and more video are encoded with H.264/AVC currently. For video temporal segmentation tasks, due to the use of some new tools in H.264/AVC, some of the features used to detect scene change in the MPEG compressed domain are not available. The major difficulty is introduced by the intra mode prediction that uses the spatial prediction information, which makes it very difficult to reconstruct the DC images that are used by a lot of previous MPEG domain works. Meanwhile, there will be more intra mode macroblocks in H.264/AVC bitstreams than MPEG bitstreams. We encode the sequence ANNI009 by MPEG-4 and H.264/AVC respectively. The macroblock types of the 823rd frame are shown in figure 2. Even there is very little motion in this frame, more intra mode macroblocks are encoded by H.264/AVC because of the smooth spatial character. The residuals may have the very low energy when using the intra prediction in H.264/AVC. If the ratio of intra mode macroblocks is used as the main feature to detect a scene change, this frame will be falsely declared in H.264/AVC format. Therefore, the ratio of intra mode macroblocks in a frame does not offer a good clue to detect the scene change for H.264/AVC bitstreams.



(a) (b)

**Figure 2 Macroblock types of the 823rd frame in ANNI009 (a) MPEG4 (b) H.264**

To solve the problem of the inaccuracy in scene change detection methods based on the ratio of intra mode macroblocks, three statistical features (BIM, BPM and the number of skip mode macroblocks) are extracted only from the variable length decoding process. Instead of using the count of intra mode and inter mode in a frame, we just need to count the number of bits that intra mode macroblocks and inter mode macroblocks are used, and the number of the macroblocks in skip mode. When a scene change occurs, there will be little motions and most of the macroblocks will be coded by intra mode, so the bits used for intra mode macroblocks will be high and the bits used for inter mode macroblocks will be low. Meanwhile, the ratio of skip mode macroblocks will also be low. Although the ratio of intra mode macroblocks in H.264/AVC bitstreams is much higher than that in MPEG bitstreams, the bit counts for these extra intra mode macroblocks are very low because these macroblocks have much lower energy. If we would use the number of bits ratio instead of the number ratio of intra mode macroblocks, it should be much more accurate to detect scene change.

## 3. Proposed algorithm

From the discussions in section II, the *IPR* is used as our main feature of the scene change detection algorithm.

$$IPR_i = \frac{BIM_i}{BPM_i + \varepsilon} \tag{1}$$

Where *BIM* represents the number of bits used to encode the intra mode macroblocks in the *i*th frame. *BPM* represents the number of bits used to encode the inter mode (including skip mode) macroblocks in the *i*th frame. $\varepsilon$ is a non-zero constant used to prevent *BIM* from divided by zero in eq.(1).

Figure 3 shows the $IPR_i$ value at the *i*th frame extracted from the sequence ANNI009.

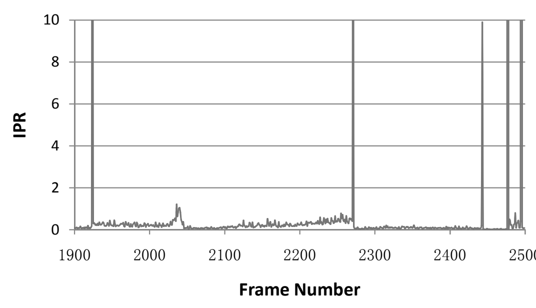



**Figure 3 IPR value for frame 1900~2500 of ANNI009**

As seen from Figure 3, there is a very high value of $IPR_i$ when an abrupt change occurs and the value around that frame is very low. There are a set of frames that have higher values of $IPR_i$ when there is a gradual change.

From the observations above, the algorithm can be divided in two progresses to detect the abrupt change and the gradual change respectively.

### 3.1. Abrupt Change Detection

For abrupt change, a sliding window whose size is (2*M+1) is used to examine the local maximum value. Let $SKP_i$ be the percent of skip mode macroblocks at the ith frame. A abrupt scene change is declared at the ith frame if (1) $IPR_i$ is larger than or equal to the sum of $IPR_j$ ($j$= $i$-M, …, $i$-1, $i$+1, …, $i$+M) and (2) $SKP_i$ is less than or equal to the minimum of the $SKP_j$ ($j$=$i$-M, …, $i$-1, $i$+1, …, $i$+M) in the slide window and $\varphi$. The parameter $\varphi$ is a predefined variable to prevent false alarm and its value is set to 5% experimentally. Meanwhile, the parameter M is set to 5 in our experiment. The conditions can be denoted as follows:

$$IPR_i \geq \left( \sum_{k=1}^{\mathrm{M}} IPR_{i-k} + \sum_{k=1}^{\mathrm{M}} IPR_{i+k} \right) \tag{2}$$

$$SKP_i \leq \mathrm{MIN}\left(\mathrm{MIN}\left(SKP_{i-\mathrm{M}}, \cdots, SKP_{i+\mathrm{M}}\right), \varphi\right) \tag{3}$$

### 3.2. Gradual Change Detection

The gradual change is detected by a four-step process. At the beginning, we assume that there will be no gradual change just after or before an abrupt change. Thus the gradual change is detected between the two abrupt changes which are declared in the previous process.

Step 1) A median-filter is performed on $IPR_i$ and $SKP_i$ to get two new features $IPR_i'$ and $SKP_i'$ :

$$IPR_i' = \mathrm{MEDIAN}\left(IPR_{i-1}, IPR_i, IPR_{i+1}\right) \tag{4}$$

$$SKP_i' = \mathrm{MEDIAN}\left(SKP_{i-1}, SKP_i, SKP_{i+1}\right) \tag{5}$$

Step 2) Then $IPR_i'$ and $SKP_i'$ are filtered by a mean-filter to get two smoother features, $IPR_i''$ and $SKP_i''$ :

$$IPR_i'' = \frac{1}{2 \times M + 1} \times \sum_{k=-M}^{M} IPR_{i+k}' \tag{6}$$

$$SKP_i'' = \frac{1}{2 \times M + 1} \times \sum_{k=-M}^{M} SKP_{i+k}' \tag{7}$$

$M$ is set to 5 in our experiment.

The first two steps efficiently reduce the noise effects.

Step 3) $IPR_i''$ and $SKP_i''$ are used to decide whether there is a gradual change.

$$\sum_{k=-1}^{1} IPR_{i+k}'' \geq 3 \times \mathrm{MAX}\left(IPR_{i-20}'', \cdots, IPR_{i-10}''\right) \tag{8}$$

$$SKP_i'' \leq \mathrm{MIN}\left(\mathrm{MIN}\left(SKP_{i-20}'', \cdots, SKP_{i-10}''\right), \phi\right) \tag{9}$$

$\phi$ is set to 10% in our experiment.

The ith frame which is satisfied by (8) and (9) is declared as a gradual change candidate.

Step 4) A group algorithm is used to group the selected frames which form a true gradual change. If the adjacent selected frame numbers are less than a distance $\alpha$ , then these frames are grouped together, and the number of the grouped frames must larger than $\beta$ :

$$GC = \left\{c_i, \cdots, c_j\right\}, \mathrm{f}\left(c_{k+1}\right) - \mathrm{f}\left(c_k\right) \leq \alpha, \forall k = i, \cdots, j-1 \tag{10}$$

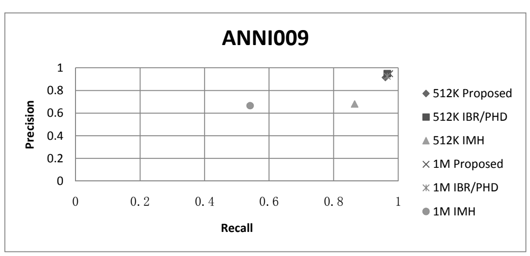
$$\mathrm{SIZE}\left(GC\right) \geq \beta \tag{11}$$

The values of $\alpha$ and $\beta$ are set to be 3 and 5.

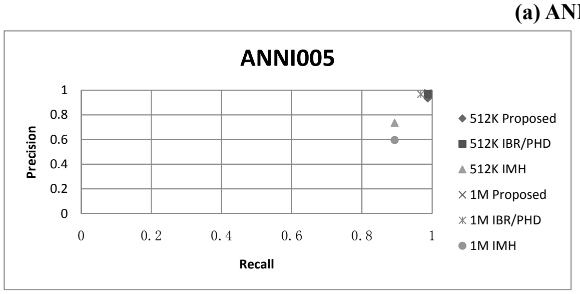
## 4. Simulation results

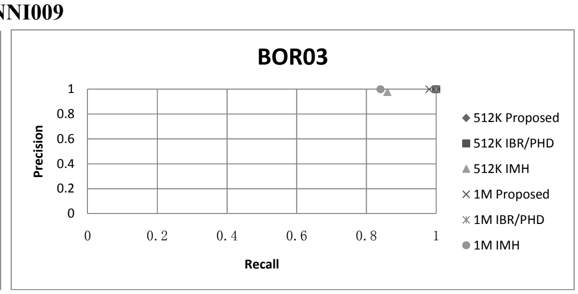
In order to evaluate the proposed algorithm, we use the sequences from TRECVID-2001 video data set. This video collection consists of many documentary style videos and it is used for many video analysis tasks. The sequences are re-encoded with H.264/AVC reference software JM8.6 at two types of bit rate: 512kbps and 1Mbps.

Two conventional performance measures are used for scene change detection algorithm; recall and precision which are defined as follows:



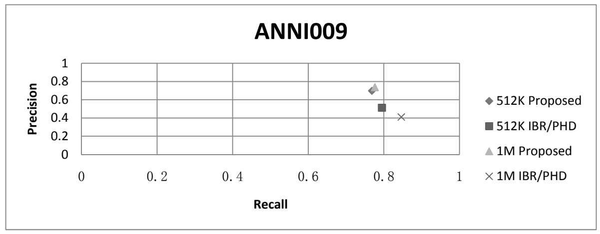

**(a) ANNI009**



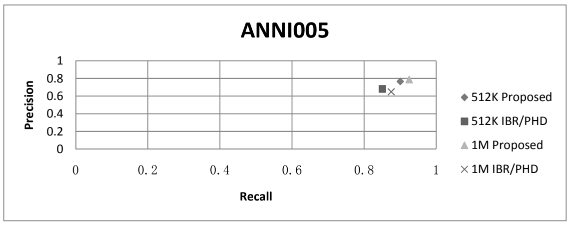




**(b) ANNI005** **(c) BOR03**

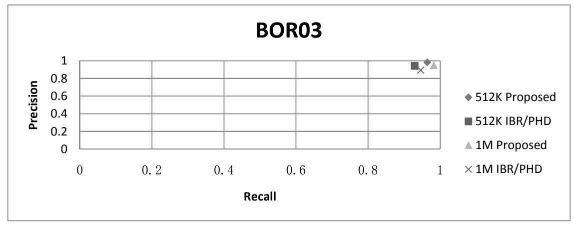
**Figure 4 The precision and recall values on detection abrupt change of the three algorithms**




**(a) ANNI009**




**(b) ANNI005**




**(c) BOR03**

**Figure 5 The precision and recall values on detection abrupt and gradual change of the two algorithms**

$$\text{Recall} = \frac{N_c}{N_c + N_m}, \quad \text{Precision} = \frac{N_c}{N_c + N_f}$$

Where $N_c$ is the number of correct detections, $N_f$ is the number of false detections, and $N_m$ is the number of miss detections.

We compare our method with Schoffmann's method in [3] (PHD to detect the abrupt scene changes and IBR to detect the gradual scene changes) and Zeng's method in [6] (IMH to detect the abrupt scene changes). Since the IMH method can only detect abrupt scene change, we only compare the abrupt change detection among the three algorithms and compare both types of change detection between the proposed method and PHD/IBR.

The result of the abrupt change detection can be seen in Figure 4. The precision and recall values acquired from the proposed method and PHD have higher values than that of the acquired from the IMH method. Because the L1-Distance based partition histograms must be calculated in PHD process, it is more complex than our proposed algorithm.

The result of the abrupt and the gradual changes detection can be seen in Figure 5. As seen in these figures, the PHD/IBR method usually has a lower precision value. The reason as discussed in section II is that the only use of the ratio of intra-coded macroblocks will result in a lot of false detections in H.264/AVC compressed domain.

**Table 1 Scene change detection time in relation to decoding time of different sequences**

| Sequence | Scene change detection time in relation to decoding time |
|---|---|
| ANNI009 | 9.46% |
| ANNI005 | 9.83% |
| BOR03 | 9.11% |

Since the bit allocation for both intra and inter mode macroblocks and the number of skip mode macroblocks can be extracted very easily just from the first step in H.264/AVC decoding process, the complexity of the proposed algorithm is very low. Table 1 shows our scene change detection time in relation to the whole H.264/AVC decoding time and it costs less than 10% of the decoding time.

## 5. Conclusions

In this paper, a novel scene change detection algorithm in H.264/AVC compressed domain is proposed. There are two advantages in the algorithm: One is that we use the bit allocation for intra mode macroblocks instead of the number of intra mode in a frame to avoid many false alarms, and the other is that the algorithm based on an adaptive threshold can detect scene change more efficiently than the algorithms based on the fixed threshold. The results of experiment have shown that the proposed algorithm is robust for both the abrupt and the gradual scene change detections. Currently, we are exploring a robust algorithm for the scene change detection in H.264/AVC bitstreams with B frames.

## References


[1] I. Koprinska, S. Carrato(2001). "Temporal video segmentation: a survey". *Signal Processing:Image Communication*. Vol. 16, No. 5, pp. 477-500.

[2] T. Wiegand, G. J. Sullivan, G. Bjøntegaard and A. Luthra(2003). "Overview of the H.264/AVC video coding standard". *IEEE Transactions on Circuits and Systems for Video Technology*. Vol. 13, No. 7, pp. 560-576.

[3] S. M. Kim, J. Byun, C. Won(2005). "A scene change detection in H.264/AVC compression domain". *Proceedings of PCM05, LNCS*. Vol. 3768, pp. 1072–1082.

[4] B. Damghanian, M. R. Hashemi, M. K. Akbari(2006). "A novel fade detection algorithm on H.264/AVC compressed domain". *Lecture Notes in Computer Science*, Vol. 4319, pp. 1159-1167.

[5] W. Zeng, W. Gao(2005). "Shot change detection on H.264/AVC compressed video". *IEEE International Symposium on Circuits and Systems*. Vol. 4, pp. 3459–3462.

[6] S. De Bruyne, W. De Neve, K. De Wolf, D. De Schrijver, P. Verhoeve, R. Van de Walle(2007). "Temporal video segmentation on H.264/AVC compressed bitstreams".

*Lecture Notes in Computer Science*. Vol. 4810, pp. 177-186.

[7] K. Schoffmann, and L. Boszormenyi(2008). "Fast segmentation of H.264/AVC bitstreams for on-demand video summarization". *Lecture Notes in Computer Science*. Vol. 4903, pp. 265-276.

[8] A. Dimou, O. Nemethova, M. Rupp(2005). "Scene change detection for H.264 using dynamic threshold techniques". *Proceedings of 5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Service.*