An Effective and Fast Scene Change Detection Algorithm for MPEG Compressed Videos

Z. Li[1], J. Jiang[1,2], G. Xiao[1], and H. Fang[2]

[1] Faculty of Informatics & Computing, Southwest China University, Chongqin, China
j.jiang1@bradford.ac.uk, g.xiao@swcu.edu.cn
[2] Department of EIMC, University of Bradford, UK

Abstract. In this paper, we propose an effective and fast scene change detection algorithm directly in MPEG compressed domain. The proposed scene change detection exploits the MPEG motion estimation and compensation scheme by examining the prediction status for each macro-block inside B frames and P frames. As a result, locating both abrupt and dissolved scene changes is operated by a sequence of comparison tests, and no feature extraction or histogram differentiation is needed. Therefore, the proposed algorithm can operate in compressed domain, and suitable for real-time implementations. Extensive experiments illustrate that the proposed algorithm achieves up to 94% precision for abrupt scene change detection and 100% for gradual scene change detection. In comparison with similar existing techniques, the proposed algorithm achieves superiority measured by recall and precision rates.

1 Introduction

Detection of scene changes plays important roles in video processing with many applications ranging from video indexing and video summarization to object tracking and video content management. Over the last three decades, scene change detection has been widely studied and researched. As a result, many scene change detection techniques have been proposed and published in the literature. For our convenience of surveying existing research in this subject area, all these algorithms and techniques can be broadly classified as operating on decompressed data (pixel domain), or working directly on the compressed data (compressed domain).

In pixel domain, the major techniques are based on pixels, histogram comparison and edge difference examinations [1],[2],[15],[16],[17],[18]. Seung Hoon Han [17] proposed an algorithm combining Bayesian and structural information to detect scene changes. JungHwan et. Al. [16] developed a content-based scene change detection algorithm, which computes background difference between frames, and use background tracking to handle various camera motions.

Recent trends focus on developing scene change detection algorithms directly in compressed domain, especially corresponding to MPEG compressed videos[1, 3-14]. Scene change detection algorithms operating on compressed data may use MB and motion type information for detecting transitions. Smeaton et. Al [4] developed such a technique to support content-based navigation and browsing through digital video

A. Campilho and M. Kamel (Eds.): ICIAR 2006, LNCS 4141, pp. 206 – 214, 2006.

archives. To reduce the decoding operation to its minimum, Lelescu and Schonfeld [5] have proposed a real-time approach for detecting scene changes based on statistical sequential analysis directly on compressed video bitstreams. Dulaverakis et al [7] have proposed an approach for the segmentation of MPEG compressed video, which relies on the analysis and combination of various types of video features derived from motion information in compressed domain. Edmundo et al [9] have proposed an edges and luminance based algorithm to detect scene changes. At present, MPEG compression schemes have been widely studied, and its motion estimation and compensation has been exploited to detect scene changes by using MB type information and motion vectors [1,12~14]. Fernando et al [12] proposed that abrupt scene change could be detected by computing the number of interpolated macroblocks and the number of backward macroblocks inside B frames. In an approach that can detect both abrupt and gradual scene changes with MB type information[1], the authors explore the number of intra macroblocks and backward macroblocks to detect absurpt scene changes, and the number of interpolated macroblocks to detect gradual scene changes. Johngho et al [14] explored each macroblock in a B-frame is compared with the type of corresponding macroblock (i.e. the macroblock in the same position) of the previous B-frame to detect abrupt scene changes.

In this paper, via focusing on MB type information and coding modes, we propose a fast scene change detection method, which features that threshold selection and comparisons are made adaptive to the input video content. By using the number of intra-coded macroblocks inside P frame and the number of interpolated macroblocks inside B frames, the abrupt scene changes and gradual changes can be automatically detected. In addition, the number of backward-predicted macroblock and forward-predicted macroblocks inside B frames are also used as an enhancement to improve the precision.

The rest of the paper is structured as follows. While section 2 describes the algorithm design, section 3 reports some experiments with real video films to derive empirically decided parameters, and section 4 reports experimental results in comparison with the existing algorithms, and section V provides conclusions and discussion on possible future work.

2 The Proposed Algorithm Design

To achieve effective and efficient motion estimation and compensation inside the video compression scheme, MPEG arranged video sequences into group of pictures (GoP), for which the structure of such arrangement can be illustrated in Figure-1. As seen, the MPEG video structure has the feature that there exist two B-frames between every pair of I-P frames or P-P frames. To make it more convenient in describing the proposed algorithm, we follow the work reported in [1] to label the first B-frame as B_f (front-B), and the second B-frame as B_r (rear-B). As a result, each GoP can be further divided into two sub-groups, which are IB_fB_r and PB_fB_r, and thus the proposed algorithm can also be designed in terms of the two sub-groups.

Essentially, the proposed algorithm follows the principle that the MPEG properties and its embedded motion estimation and compensation scheme should be further exploited for scene change detection[1,2], where the status of such motion estimation

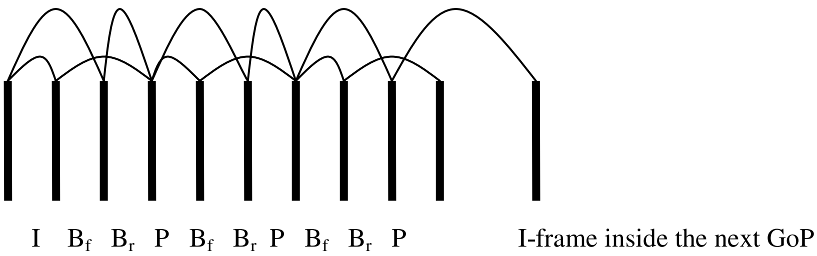

Fig. 1. Illustration of MPEG video structure for motion estimation & compensation

and compensation for each macro-block can be recorded to design such scene change detection algorithm. To this end, Pei and Chou [1] proposed a simple MB-type based scene change detection algorithm for both abrupt scene changes and gradual scene changes. The principle they applied is to monitor the number of MBs in intra-coding mode inside P-frames. Whenever the number of intra-coded MBs is above a pre-defined threshold, two separate operations for scene change detection are activated for abrupt changes and gradual changes respectively. For abrupt changes, their detection is based on one of the three motion estimation and compensation forms in MPEG videos as illustrated in Figure-2. For gradual changes, their detection is based on conditions. One is that a significant number of MBs inside P-frames are intra-coded, indicating significant change of content, and the other is that a dominant number of MBs inside B-frames are interpolative motion compensated. Detailed analysis reveals that there exist a range of weaknesses in such an algorithm, which include: (i) the scene change detection is dependent on four fixed and pre-defined thresholds; (ii) abrupt change detection and gradual scene change detection are two separate operation procedures; and (iii) the performance is low in terms of precision in detecting scene changes. To this end, we propose to: (i) introduce an adaptive mechanism in selecting all the thresholds, and make them adaptive to the input video content; (ii) combine abrupt change detection and gradual change detection into an integrated algorithm; and (iii) add a consecutive detection window to improve the gradual change detection.

Given the fact that specific motion estimation and compensation status for each macro-block is indicated by the number of bi-directionally predicted blocks for B-frames, and intra-coded macroblocks for P-frames, we define a range of essential variables as follows:

N_i: the number of intra-coded MBs inside P-frames

N: the total number of macroblocks inside each video frame;

N_{fb}: the number of both forward and backward predicted macroblocks inside each B-frame;

N_b: the number of backward predicted macroblocks inside each B-frame, indicating that this macroblock is only predicted in backward mode;

N_f the number o forward predicted macroblocks inside each B-frame.

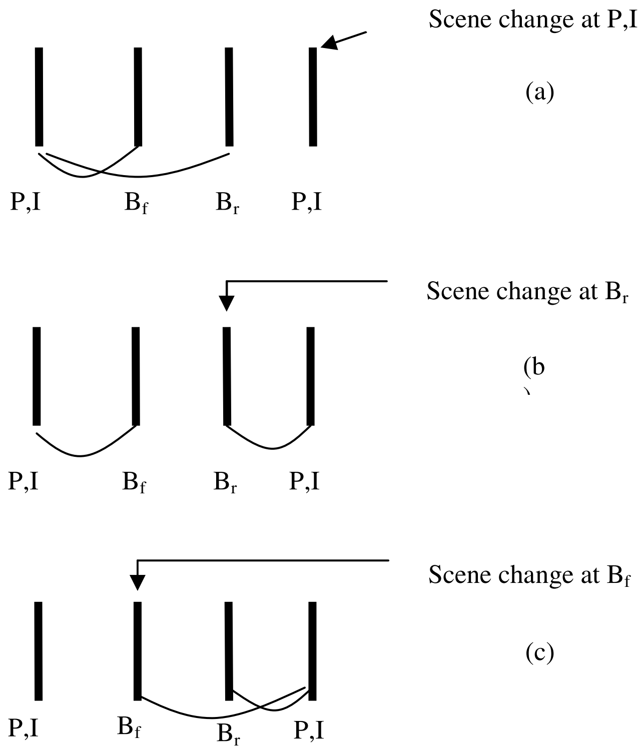

Fig. 2. Illustration of abrupt scene change detection

To decide whether the scene change detection should be activated or not, the following test is conducted for every P-frame, where the proportion of intra-coded MBs is examined with respect to a predefined threshold:

$$\frac{N_i}{N} > T_1 \tag{1}$$

Where T_1 stands for a threshold, which is to be determined empirically.

From (1), a number of observations and hence analysis can be made. Firstly, the essence of the ratio between N_i and N reflect the proportion of predicted macroblocks inside this P-frame. When the ratio is smaller than the threshold, it indicates that most of the macroblocks can be motion compensated by its reference frame, and hence a significant extent of correlation between this P-frame and it reference frame can be established. Therefore, it is not likely that there exists any scene change around this P-frame. As a result, we should carry on examining the next P-frame. Secondly, if the condition represented by (1) is satisfied, it should indicate that most of the macroblocks are not well compensated by the reference frame. Therefore, it is likely that there may exist a scene change in the neighbourhood of this P-frame. As a result, further confirmation of such scene change needs to be confirmed by examining the

B-frames to find out: (i) whether such scene change is abrupt or gradual; (ii) its exact location of such scene changes.

As illustrated in Figure-1, we have front-B and rear-B to be examined at this stage. For all B-frames, there are three possibilities for their specific process of motion estimation and compensation, which include: forward prediction, backward prediction, and bi-directional prediction. To detect the situation given in Figure-2, we propose a

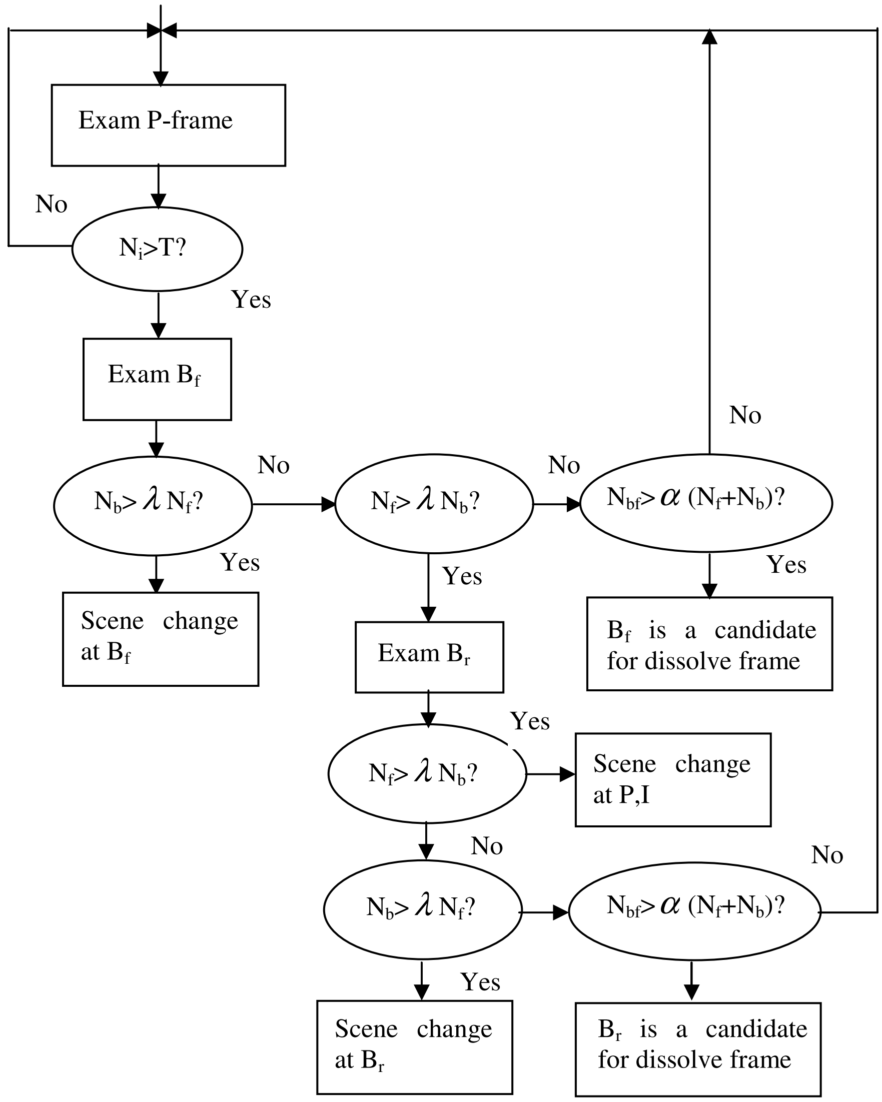

Fig. 3. Overview of the proposed scene change detection algorithm

test, where its threshold is adaptive to the input video content and its motion estimation and compensation process. Therefore, to examine the front-B frame, we have:

$$N_b > \lambda N_f \tag{2}$$

Where λ is a parameter controlling the balance between the backward prediction and forward prediction.

Satisfaction of (2) indicates that backward predicted MBs overwhelm the forward predicted MBs. As a result, the possible scene change can be further confirmed, and located at B_f as shown at the bottom of Figure-2. If the test is negative, we need to find out whether forward predicted MBs overwhelm the backward predicted MBs, in order to verify whether any of the two forms given in (a) and (b) of Figure-2 can be detected. Such condition can be tested by similar equation as illustrated below:

$$N_f > \lambda N_b \tag{3}$$

Satisfaction of the above test confirms either (a) or (b) in Figure-2 depending on the result of examining the rear-B frame. In this circumstance, we firstly repeat the test given in (3) on the rear-B frame. Satisfaction of such test will confirm the case of (a) in Figure-2 and hence a scene change can be detected at the next P or I frame. If the test is negative, we check if the backward predicted MBs overwhelm the forward predicted MBs inside the rear-B frame by (2). Positive test indicates that scene change is at the B_r frame as given in (b) of Figure-2. If not, we need to check if B_r frame is a candidate for dissolve frame (gradual change) by the following test:

$$N_{bf} > \alpha(N_f + N_b) \tag{4}$$

Where α is another parameter indicating the dominance of bi-directional prediction of MBs inside B-frames.

Similarly, negative test of (3) on the front-B frame will prompt the test (4) being applied to B_f to see if it is a candidate for a dissolve frame.

In summary, the proposed algorithm can be overviewed in Figure-3.

3 Algorithm Evaluations

To evaluate the performance of the proposed scene change detection algorithm, a database of video sequences has been obtained from recording of TV programmes, including documents, news, sports, and films. The database represents a total of 114 abrupt scene changes and 22 dissolves in 15 clips of videos. The selected sequences are complex with extensive graphical effects. Videos were captured at a rate of 30 frames/s, 640×480 pixel resolution, and stored in PPM format. Details of the test video clips are described in Table-1.

For benchmarking purposes, we selected the algorithm reported in [1] as a representation of the existing techniques to provide a comparison for the evaluation of the proposed algorithm. Performance measurements are implemented with the well-known recall and precision figures [16], which are defined as follows.

$$\text{Recall}=\frac{\det ects}{(\det ects+MD)} \tag{5}$$

$$\text{Precision}=\frac{\det ects}{(\det ects+FA)} \tag{6}$$

Table 1. Description of video clip test set

Video sequences	Number of abrupt scene changes
Clip-1	24
Clip-2	15
Clip-3	13
Clip-4	21
Clip-5	19
Clip-6	22
Clip-7	86~107;160~179
Clip-8	135~165;540~574
Clip-9	80~115;230~258
Clip-10	6~101,128~144;315~325
Clip-11	5~16;209~248;358~365
Clip-12	27~42;67~80;140~181
Clip-13	70~88;115~130;190~289
Clip-14	132~175;224~278
Clip-15	23~44;158~189
total	Abrupt=136; dissolved=22

Where *detects* stands for the correctly detected boundaries, while MD and FA denote missed detections and false alarms, respectively. In other terms, at fixed parameters, recall measures the ratio between right detected scene changes and total scene changes in a video, while precision measures the ratio between the right detected scene changes and the total scene changes detected by algorithm.

The final results of the evaluations are given in Tables 2.

From Table 2, it can be seen that the proposed method achieves on average a 90.38% recall rate with a 75.49% precision rate for both abrupt and gradual scene changes. In comparison with the existing counterparts [1] given in Table-2 as the benchmark, the proposed method achieves superior performances in terms of both recall and precision rates. Additional advantages with the proposed method can be highlighted as: (i) integrated algorithm for both abrupt scene change and dissolve detection; (ii) directly operates in compressed domain and thus suitable for real-time implementation; and (iii) only one threshold and two parameters are required for scene change detection and yet the detection mechanism is made adaptive to the input video content, or the performance of motion estimation and compensation embedded inside MPEG techniques.

Table 2. Summary of experimental results

	The proposed algorithm		The benchmark	
Clips	Recall	Precision	Recall	Precision
1	87	90	56	85
2	92	92	72	80
3	96	86	76	78
4	89	85	65	75
5	90	94	84	81
6	85	92	80	89
7	100	100	100	66.7
8	100	66.7	100	50
9	50	50	50	33.3
10	100	75	66.7	50
11	100	60	100	43
12	66.7	75	100	43
13	100	50	66.7	50
14	100	66.7	100	40
15	100	50	100	50
Average	90.38	75.49	81.09	60.93

5 Conclusions

In this paper, we described an integrated algorithm, which is capable of directly and effectively detect scene changes in MPEG compressed videos. The proposed algorithm works in compressed domain exploiting existing MPEG motion estimation and compensation mechanisms. Therefore, it achieves significant advantages in terms of processing speed and algorithm complexity, and thus suitable for fast and real-time implementations. While the proposed algorithm can save a lot of computation cost, extensive experiments support that the proposed algorithm also achieves superior performances over the existing counterparts.

Finally, the authors wish to acknowledge the financial support under EU IST FP-6 Research Programme as funded for the integrated project: LIVE (Contract No. IST-4-027312).

References:

1. Aidan Totterdell, "An algorithm for detecting and classifying scene breaks in MPEG video bit streams", Technical report , No.5, September 1998;
2. Ramin Zabih, Justin Miller and Kevin Mai, "A feature-based algorithm for detecting and classifying production effects", Multimedia Systems ,Vol.6,No.2, 1999 , pp.119-128;
3. Georgios Akrivas,Nikolaos.D.Doulamis,Anastasios.D.Doulamis and Stefanos.D.Kollias, " Scene detection methods for MPEG-encoded video signals", Proceeding of the MELECON 2000 Mediterranean Electrotechnical Conference, Nicosia, Cyprus, May 2000;

4. A. Smeaton et al, "An evaluation of alternative techniques for automatic detection of shot boundaries in digital video", in Proc. Irish Machine Vision and Image Processing Conference (IMVIP'99), Dublin, Ireland, 8-9 Sep. 1999, pp.45-62;
5. Dan Lelescu and Dan Schonfeld, " Statistical sequential analysis for real-time video scene change detection on compressed multimedia bitstream", Multimedia, IEEE Transactions on, Vol.5, No.1, March 2003, pp.106-117;
6. Jung-Rim Kim, Sungjoo Suh, and Sanghoon Sull, "Fast scene change detection for personal video recorder", IEEE Transactions on Consumer Electronics , Vol.49,No.3, August 2003, pp.683-688;
7. C.Dulaverakis, S.Vagionitis, M.Zervakis and E.Petrakis, "Adaptive methods for motion characterization and segmentation of MPEG compressed frame sequences", International Conference on Image Analysis and Recognition(ICIAR 2004), Porto, Portugal, September 29-October 1, 2004, pp.310-317;
8. S.Porter, M.Mirmehdi, B.Thosma, "Temporal video segmentation and classification of edit effects", Image and Vision Computing, Vol.21, No.13-14, December 2003, pp.1098-1106;
9. Edmundo Saez, Jose I.Benavides and Nicolas Guil, "Reliabele time scene change detection in MPEG compressed video", ICME 2004, 27-30 June 2004, Taipei, pp.567-570;
10. Jesus Bescos , "Real-time shot change detection over online MPEG-2 video", Circuits and Systems for Video Technology, IEEE Transactions on ,Vol.14, No.4, April 2004, pp.475-484;
11. Serkan Kiranyaz, Kerem Caglar, Bogdan Cramariuc and Moncef Gabbouj, "Unsupervised scene change detection techniques in features domain via clustering and elimination", Proc. of Tyrrhenian International Workshop on Digital Communications, Capri, Italy, September 8th - 11th, 2002;
12. W.A.C.Fernando, C.N.Canagarajah and D.R.Bull, "Scene change detection algorithms for content-based video indexing and retrival", Electronics&communication engineering journal, June 2001, pp.117-126;
13. Soo-Chang Pei and Yu-Zuong Chou, "Novel error concealment method with adaptive prediction to the abrupt and gradual scene changes", IEEE transactions on multimedia, Vol. 6, No.1, February 2004, pp.158-173.
14. Johngho Nang, Seungwook Hong and Youngin Ihm, " An efficient video segmentation scheme for MPEG video stream using macroblock information", Proceedings of the 7th ACM International Conference on Multimedia '99, October 30 - November 5, 1999, Orlando, FL, USA. ACM, 1999, Vol.1, pp.23-26;
15. Shu-Ching Chen, Mei-Ling Shyu, Cheng-Cui Zhang and R.L.Kashyap, "Video scene change detection method using unsuper segmentation and object tracking" , IEEE International Conference on Multimedia and Expo(ICME), 2001, pp.57-60;
16. JungHwan Oh**, Kien A.Hua**, Ning Liang **, "A content-based scene change detection and classification technique using background tracking" , Proc. of IS&T/SPIE Conference on Multimedia Computing and Networking 2000, San Jose CA, January 2000, pp.254-265;
17. Seung Hoon Han, "Shot detection combining bayesian and structural information", Storage and Retrieval for Media Databases 2001, Vol. 4315, December 2001, pp509-516;
18. H.B.Lu and Y.J.Zhang, "Detecting abrupt scene change using neural network", Visual Information and Information Systems, Third International Conference, VISUAL '99, Amsterdam, The Netherlands, June 2-4, 1999, pp.291-298.