Efficient MPEG Compressed Video Analysis Using Macroblock Type Information

Soo-Chang Pei, *Senior Member, IEEE*, and Yu-Zuong Chou

***Abstract*— Efficient indexing methods are required to handle the rapidly increasing amount of visual information within video databases. Video analysis that partitions the video into clips or extracts interesting frames is an important preprocessing step for video indexing. In this paper, we develop a novel method for video analysis using the macroblock (MB) type information of MPEG compressed video bitstreams. This method exploits the comparison operations performed in the motion estimation procedure, which results in specific characteristics of the MB type information when scene changes occur or some special effects are applied. Only a simple analysis on MB types of frames is needed to achieve very fast scene change, gradual transition, flashlight, and caption detection. The advantages of this novel approach are its direct extraction from the MPEG bitstreams after VLC decoding, very low complexity analysis, frame-based detection accuracy and high sensitivity.**

***Index Terms*— Compressed domain video analysis, flashlight and caption detection, scene change, video indexing.**

I. Introduction

IN THE MODERN world, the amount of visual information is growing larger and wider just like the number of cable TV channels we can choose from nowadays. Amidst this huge amount of visual information, searching for the video sections in which we are interested is a very difficult task. Traditionally, all we can do is fast-forward the video and keep our eyes on the screen for this purpose. This is very time-consuming and labor-intensive. Therefore, an efficient video browsing and retrieval system is required to improve the viewing experience. A fundamental and important processing step to accomplish this objective involves analyzing the video sequence and parsing the video into a set of clips. The most straightforward video indexing technique entails segmenting the video into temporal shots, each of which represents a continuous sequence of actions. Recently, detecting the boundaries of shots that are so-called scene changes has become an important topic and has attracted a lot of research.

Up to now, many approaches have been developed to detect scene changes in video sequences [1]–[8]. In order to lower the computation cost, more recent research focuses on performing the scene change detection directly on compressed video data instead of raw video. Generally, a scene change, or scene cut, is defined as an image content switch between two consecutive frames with different scenes. Because of the change of video content, two consecutive frames in different shots show significant differences in many characteristic features. Tonomura *et al*. [1] proposed a method to exploit the histogram difference of consecutive frames to detect scene changes. Arman *et al*. [2] developed a content-based video browser using shape and color content analyzes. In his paper, spatial moments and color histogram features are extracted for shape and color information. The previous two methods can perform well but they have the drawback of high computation load of decompressing, since video data is usually stored in the compressed domain. Therefore, some methods focus on processing directly on conventional compressed data standards, such as H.261, MPEG, etc. Most of these methods utilize the correlation of DCT coefficients between frames based on the idea that similar spatial content mathematically results in similar DCT coefficients. Arman *et al*. [3] have proposed a method to partition motion JPEG video sequences and experimental results show that the relation of DCT coefficients can be used to detect scene changes. Concerning MPEG compressed video data, Nakajima [4] has developed a fast scene change detection approach by calculating the correlation of DCT coefficients of I frames. In Nakajima's paper, he introduced the idea of spatial–temporal scaling used to improve the detection speed and lower the computation load. However, camera motion would affect the detection results and only one scene change can be detected if there are more than two scene changes existing in one group of pictures (GOP). Besides Nakajima's approach, Arman *et al*. [5] have also proposed a new method to speed up scene change detection. This new method takes the relationship between video data in spatial and frequency domains into consideration and it reduces the number of processed blocks by excluding the blocks with less representation. For example, it can only process the scene change detection on the blocks with significant high and medium frequency components that represent edges in the spatial domain. Although these fast algorithms have been proposed, camera motion problems such as panning and zooming are still not solved. Panning and zooming can totally change the spatial content of each block and, of course, the DCT coefficients. Consequently, camera motions would disturb the scene change detection because they cause a lot of false alarms. Zhang *et al*. [6] have developed an approach to exploit the information of motion vectors between consecutive P frames to detect the camera motion.

Manuscript received December 15, 1998; revised October 5, 1999. This work was supported by the National Science Council, R. O. C., under Contract NSC 89-2213-E-002-092. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Thomas R. Gardos.

The authors are with the Department of Electrical Engineering, National Taiwan University, Taipei 106, Taiwan, R.O.C. (e-mail: pei@cc.ee.ntu.edu.tw).

Publisher Item Identifier S 1520-9210(99)09968-X.

Their approach can efficiently distinguish real scene changes from those caused by camera motion.

Besides camera motion, there is another factor that can affect scene change detection accuracy. When some effects such as dissolve, fade-in, fade-out are applied to the video sequence, the boundaries of a shot do not lie sharply between two consecutive frames. It becomes a gradual transition of change from one scene to another. In this case, to be aware of this transition and to find the starting and ending points of a shot, more complicated analysis is usually needed. Zhang *et al*. [7] have proposed a multipass, twin-comparison approach to solve this problem. Their method performs a first pass to locate potential boundaries due to both abrupt scene changes and gradual transitions, then a second pass to differentiate these two types apart.

Besides scene changes, some other special events such as *Flashlight* and *Caption* may interest viewers and be applied to video browsing as key frames. Flashlights in scenes may represent the happening of highlight events or appearances of very important persons. Captions in video sequences like news or sports programs can provide important information on the underlying stories. Both of these can provide viewers the important key frame for browsing. Yeo *et al*. [8], [13] proposed a method using reduced DC images constructed from compressed MPEG video to detect the frames with flashlights and captions. Very good performance is achieved in their paper.

In this paper, we propose a novel method to exploit the comparison process in the MPEG motion estimation step that is revealed in the macroblock (MB) type information. As an abrupt scene change occurs, we observe a specific statistical feature of MB types of B frames due to the different contents of the referencing frames. For gradual scene change detection, the dominant interpolation MB's resulting from the interpolation characteristic of dissolve effect help us to distinguish gradual scene changes from fast panning sequences. Similarly, a frame containing flashlights is different from others and also affects the MB type information. A significant number of intra-coded MB's coming from the bright background of the flashlight can indicate the occurrence of the flashlight. Caption embedding would make a specific region become a scene change region. Abrupt scene change detection focusing on specific fixed marked region can perform as caption detection. Compared with other proposed methods, the extraction of MB type information from compressed video data is much easier than the conventional approaches using DCT coefficients or motion vectors. It can be obtained directly from bit streams of video sequences after only variable length decoding. Moreover, the frame-based detection accuracy of this algorithm is another important advantage. In other words, our detection method can precisely indicate which frame the scene change occurs. Last but not least, simple analysis of low computation complexity on MB type information can achieve efficient scene change, flashlight and caption detection.

II. Proposed Method

We define three subgroups of pictures (SGOP) as three sets of four frames (PBBP, PBBI and IBBP) in a GOP, which

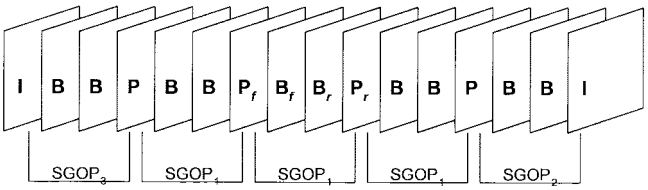


Fig. 1. Illustration of GOP structure used in the proposed method.

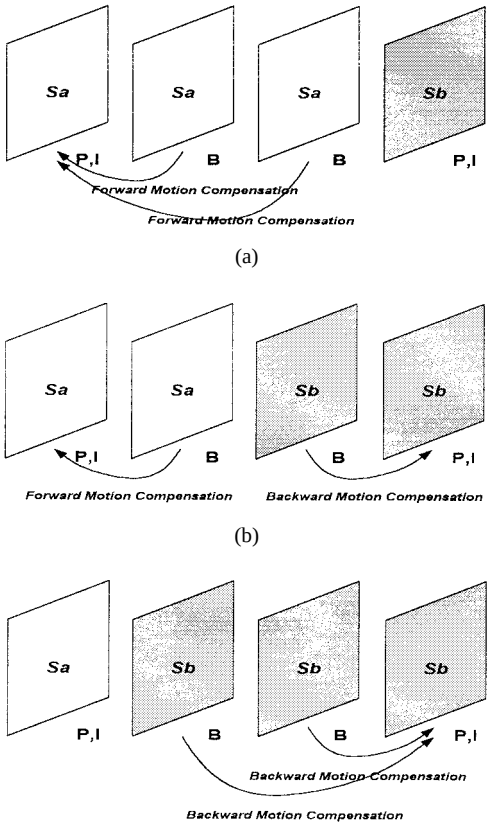


Fig. 2. Patterns of MB types in SGOP's with abrupt scene changes. (a) Scene change at P frame or I frame (SCPI). (b) Scene change at front B frame (SCFB). (c) Scene change at rear B frame (SCRB).

are represented by $SGOP_1$, $SGOP_2$ and $SGOP_3$ respectively. The frames of $SGOP_1$ are denoted as $P_f B_f B_r P_r$ (front P, front B, rear B and rear P). The structure of SGOP's in a GOP is illustrated in Fig. 1. It is noted that the MPEG encoding process is not defined, so different encoders may adopt different GOP structures. In this paper, the proposed method is developed and tested based on the specific GOP structure illustrated in Fig. 1 and the modifications of the proposed method corresponding to different GOP structures are discussed in Section IV-B.

A. Pattern of MB Types of SGOP's with Abrupt Scene Changes

In this paper, a scene change, or scene cut, is defined as an image content switch between two consecutive frames with different scenes. We classify abrupt scene changes into three types: Scene change at a P frame or an I frame (SCPI); Scene change at front B frame (SCFB); Scene change at rear B frame

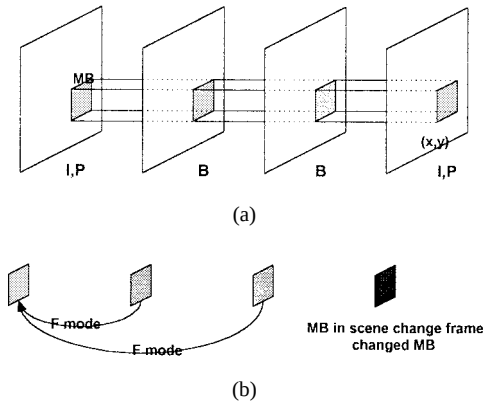


Fig. 3. (a) Illustration of MBGxy. (b) Example of pattern of MB type in the case of SCPI.

(SCRB) in a SGOP (PBBP, PBBI, IBBP). This classification covers all frames at which an abrupt scene change might occur, because the first frame in the SGOP is also the last frame in previous SGOP. Therefore, only the last three frames in the SGOP are considered for video analysis. Each type of abrupt scene changes results in a specific pattern of MB type information. In the case of SCPI, most MB's in the two B frames (B_f, B_r) are **forward motion compensated** (F *mode*) because they are much more similar to P_f than to P_r in which a scene change occurs. In addition, a significant number of MB's in P_r are **Intra-coded** (I *mode*) due to the change of the video content. Similarly, if the scene change occurs at an I frame (in $SGOP_2$), most MB's in B_f and B_r will be inclined to be F mode. In the case of SCFB, there are also many I mode MB's existing in P_r ($SGOP_1$, $SGOP_3$), but the properties of MB types of B frames are different from that of SCPI. Owing to the scene change at B_f and the great difference between the content of P_f and the following two B frames, most of the MB's in B_f and B_r will be inclined to be **Backward motion compensated** (B *mode*). It is the same case if there is an I frame ($SGOP_2$) following the two B frames instead of a P frame. In the last case, namely, SCRB, most MB's in B_f will be F mode predicted to P_f, and most MB's in B_r will be B mode predicted to P_r. Fig. 2 shows the patterns of MB type information in the cases of SCPI, SCFB and SCRB.

It is worthwhile to mention that in our method we extend these concepts to detect the scene change of a MB instead of detecting the scene change of an entire frame. We define the MB in a specific position (x,y) of each frame of a SGOP to form a MB group (MBGxy). If the pattern of the MB types in a MBGxy follows the above pattern of SCPI, SCFB or SCRB, we can say that there is a changing MB in the position (x,y) laying at P_f, B_f or B_r (illustrated in Fig. 3). This idea of scale-down can help us to detect abrupt scene change and to set up the threshold more easily. Besides, it can also be applied to the specific area of a frame instead of the entire frame for other detection purposes.

B. Pattern of MB Types in SGOP's with Gradual Scene Changes

The most common form of gradual scene change is dissolve. A dissolve operation applied from scene S_0 to scene S_n in the

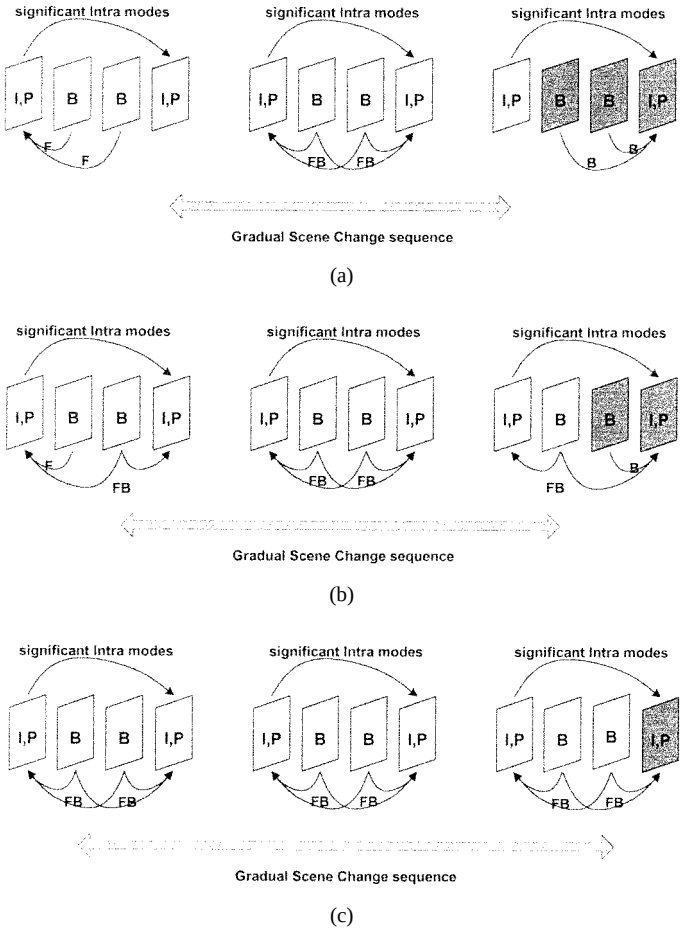


Fig. 4. Patterns of MB types with gradual scene changes. (a) Starting frame at P or I frame and ending frame at P or I frame (SFPI and EFPI). (b) Starting frame at rear B frame and ending frame at front B frame (SFRB and EFFB). (c) Starting Frame at front B frame and ending frame at rear B frame (SFFB and EFRB).

transition sequence, whose length is $n+1$, can be represented as follows:

$$S_x = \frac{x}{n} S_n + \frac{n-x}{n} S_0, \quad 0 \le x \le n. \tag{1}$$

If $S_0 = 0$, this special case of dissolve effect is also called fade-in and $S_n = 0$ is called fade-out. An example of dissolve is demonstrated in Fig. 9(a). From (1), the luminance of B frames in $SGOP_1$ operated by the dissolve effect can be approximately written as

$$B_f \approx \frac{2}{3} P_f + \frac{1}{3} P_r, \quad B_r \approx \frac{1}{3} P_f + \frac{2}{3} P_r. \tag{2}$$

When bidirectional motion compensation is applied to B_f and B_r, three probably used compensation modes (F, B, FB) are taken into consideration and the compensated frames of these modes (F_F, F_B, F_{FB}) can be written as follows:

$$F_F = P_f, \quad F_B = P_r, \quad F_{FB} = \frac{1}{2} P_f + \frac{1}{2} P_r. \tag{3}$$

Generally, the decision of MB type of B frames is based on the variance of the compensated error. The variances for three compensation modes of B_f and B_r are calculated and

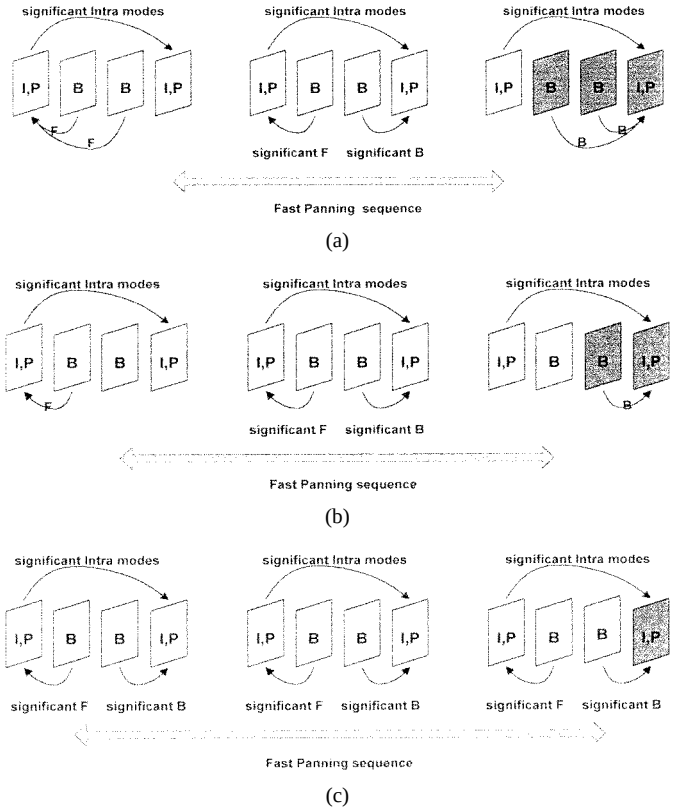


Fig. 5. Patterns of MB types with fast panning sequences. (a) Starting frame at P or I frame and ending frame at P or I frame (SFPI and EFPI). (b) Starting frame at rear B frame and ending frame at front B frame (SFRB and EFFB). (c) Starting frame at front B frame and ending frame at rear B frame (SFFB and EFRB).

formulated as follows:

$$
\begin{aligned}
Err_F(B_f) &= (B_f - F_F)^2 = \left(\frac{1}{3}P_r - \frac{1}{3}P_f\right)^2 \\
&= \frac{1}{9}(P_f - P_r)^2 \\
Err_B(B_f) &= (B_f - F_B)^2 = \left(\frac{2}{3}P_r - \frac{2}{3}P_f\right)^2 \\
&= \frac{4}{9}(P_f - P_r)^2 \\
Err_{FB}(B_f) &= (B_f - F_{FB})^2 = \left(\frac{1}{6}P_f - \frac{1}{6}P_r\right)^2 \\
&= \frac{1}{36}(P_f - P_r)^2 \\
Err_F(B_r) &= (B_r - F_F)^2 = \left(\frac{2}{3}P_r - \frac{2}{3}P_f\right)^2 \\
&= \frac{4}{9}(P_f - P_r)^2 \\
Err_B(B_r) &= (B_r - F_B)^2 = \left(\frac{1}{3}P_f - \frac{1}{3}P_r\right)^2 \\
&= \frac{1}{9}(P_f - P_r)^2 \\
Err_{FB}(B_r) &= (B_r - F_{FB})^2 = \left(\frac{1}{6}P_r - \frac{1}{6}P_f\right)^2 \\
&= \frac{1}{36}(P_f - P_r)^2. \qquad (4)
\end{aligned}
$$

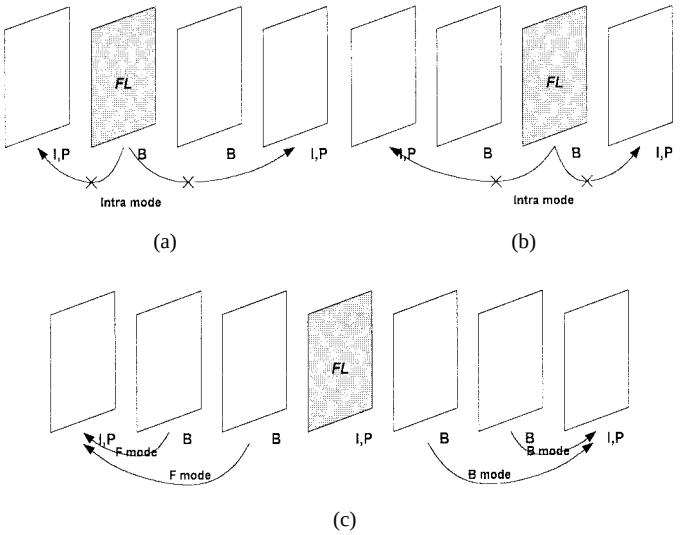
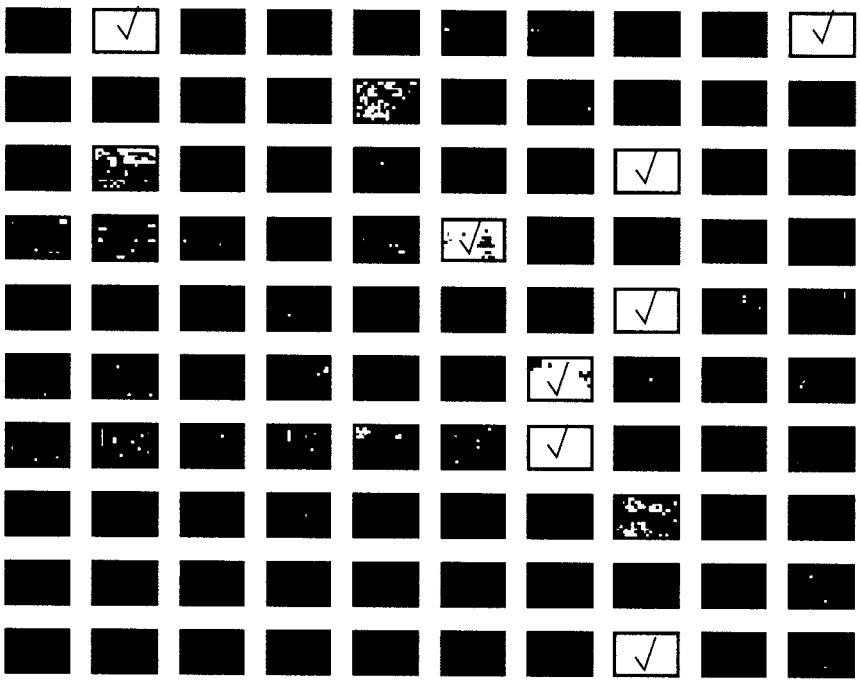


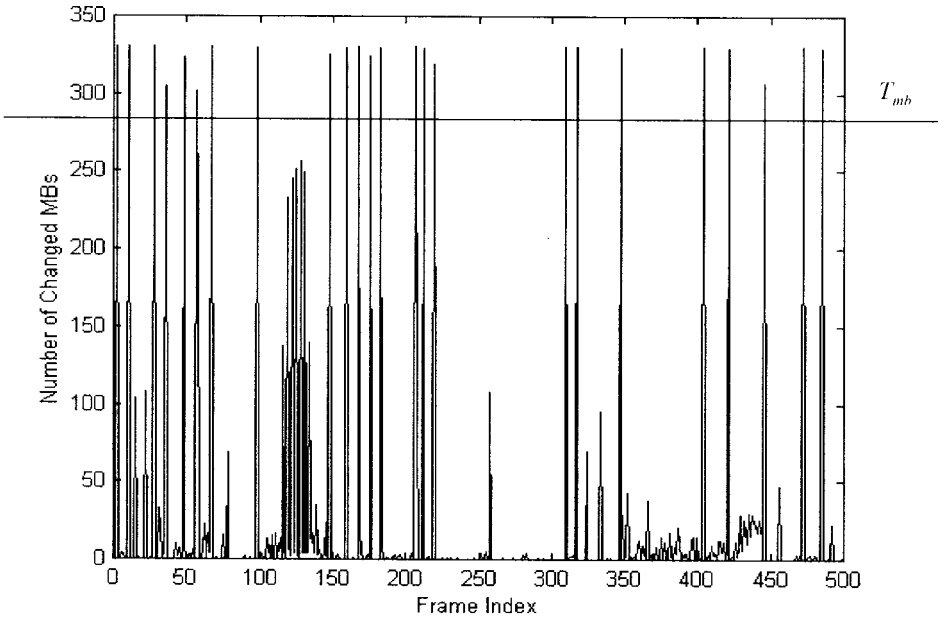
Fig. 6. Patterns of MB types in SGOP's with flashlights. (a) FLFB: flashlight lies at B_f. (b) FLRB: flashlight lies at B_r. (c) FLPI: flashlight lies at P or I.

From (4), it is found that the variance of FB compensated error is smaller than that of F or B compensated error. Consequently, B frames in dissolve video sequence are inclined to be interpolative (FB mode) motion compensated and MB's of FB mode will dominate among all MB's in B frames. It is noticeable that (3) is established under the assumption that there are no moving actions involved in the dissolve video sequence. However, if there are some moving objects existing in the dissolve video sequence, the inclination toward FB motion compensation is still valid since their influences are local.

Moreover, in the case of a gradual scene change, the main aim is to detect the starting and ending frames of the gradual transition, so we need to discuss the pattern of MB types in SGOP's including starting frames and ending frames respectively. There are three cases for the location of the starting frames defined as SFPI (starting frame lies at P_r or I frame), SFFB (starting frame lies at B_f) and SFRB (starting frame lies at B_r); and similar EFPI, EFFB and EFRB for the ending frames. In the case of SFPI, a significant number of MB's in P_r are I mode because the change of content. In addition, MB's in B_f and B_r in the SGOP containing the starting frame are inclined to be F mode because the dissolve effect has made P_r overall slightly different. In other words, the pattern of MB types of SFPI is very similar to the pattern of SCPI. However, it is noted that the inclination of B_f and B_r in SFPI toward F mode is weaker than that in SCPI because some characteristics are very similar after slight change of the dissolve effect. In the case of SFFB, most MB's in both B frames are FB mode since the frames of this SGOP form a *Complete Dissolve SGOP*, that is, the relation among four frames follows (1). In the case of SFRB, the MB's of B_f are inclined to be F mode because B_f is not affected by dissolve effect and those of B_r are inclined to be FB mode because B_r approximates the average of P_f and P_r according to (1). It is still valid that a significant number of MB's of P_r in a dissolve video sequence are I mode in both SFFB and SFRB.



(a)



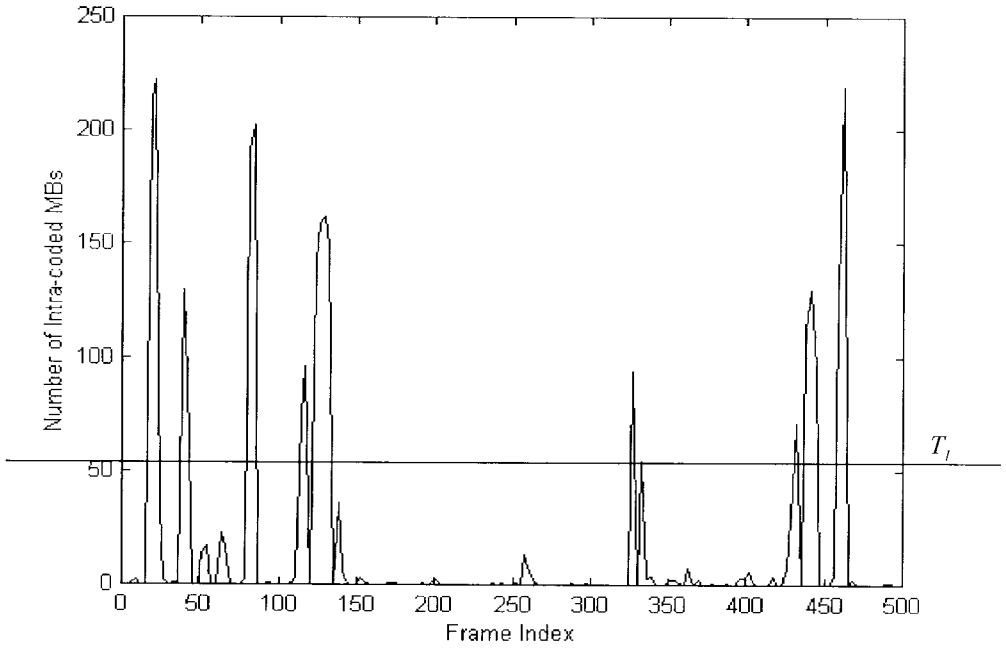
(b)

Fig. 7. (a) Locations of changing MB's in the test sequence (Frame 1 to 100), where each white dot represents a changing MB and $\surd$ locates the scene change frame. (b) Experimental results of abrupt scene change detection $(T_{\mathrm{mb}} = 280)$.

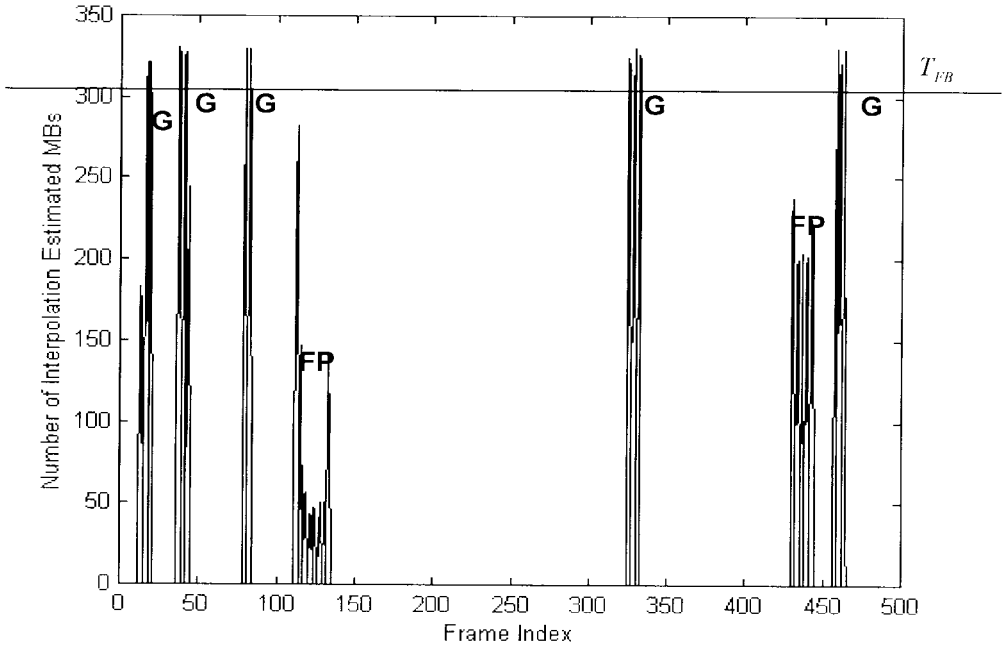
This feature is useful to indicate the potential SGOP's probably containing gradual scene changes.

The patterns of the MB types of the SGOP including ending frames are similar and symmetric to that including starting frames and both of them are shown in Fig. 4. Certainly, these three transitions shown in Fig. 4 do not include all possibilities that could happen. For example, SFPI and EFFB can also form a gradual scene change sequence. There are nine combinations but all of them are easily understood from Fig. 4. It is worthy to note that under the situation of $SGOP_2$ instead of $SGOP_1$ or $SGOP_3$, there is no clue for indicating the potential SGOP. Therefore, a $SGOP_2$ is considered a potential SGOP if the adjacent SGOP is detected as a potential SGOP in order to prevent miss detection.

A great advantage of using MB type information to detect scene changes is that they can avoid the false alarms due to large luminance difference resulting from moving objects when using some other detecting method. This advantage is



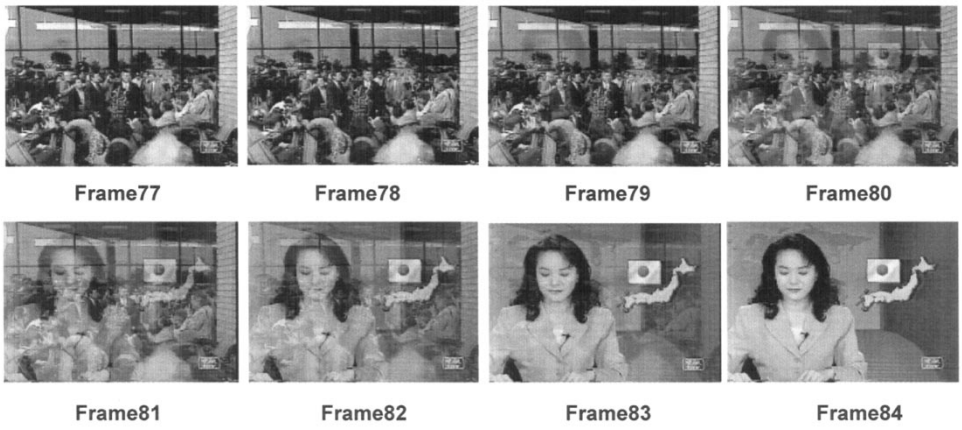
(a)



(b)

Fig. 8. Experimental results of gradual scene change detection. (a) Remaining P frames with significant I mode MB's. (b) Number of FB mode MB's in B frames in potential SGOP's (G: gradual scene change, FP: fast panning, $T_I = 50$, $T_{FB} = 300$).

still valid when dealing with the video sequence involving camera motions. Generally, zooming operation is slow for the purpose of comfortable watching. As a result, the zooming operation can be easily tracked as a special kind of movement and will not greatly affect the MB types of frames. Therefore, a video sequence involving zooming operation would be ignored by the detection process and regarded as an ordinary video sequence. It is a similar case for slow panning operation, for example, that applied to the well-known sequence 'Windmill'. Slow panning operations would be judged as overall movement of the frame and the MB type information would not show some specific characteristic. However, once the speed of panning exceeds the tracking ability of motion compensation, that is, the search window per each P frame, it confuses scene change detection and needs to be taken into consideration. Fast panning is often seen in sport video sequences, for example, tracking the football passed by the quarterback in a football game. Because the panning is too fast to be tracked,



(a)

(b)

(c)

Fig. 9. (a) Example of gradual scene changes. (b) Example of flashlight in news experiment sequence (Frame 31). (c) Three examples of embedded and removed captions.

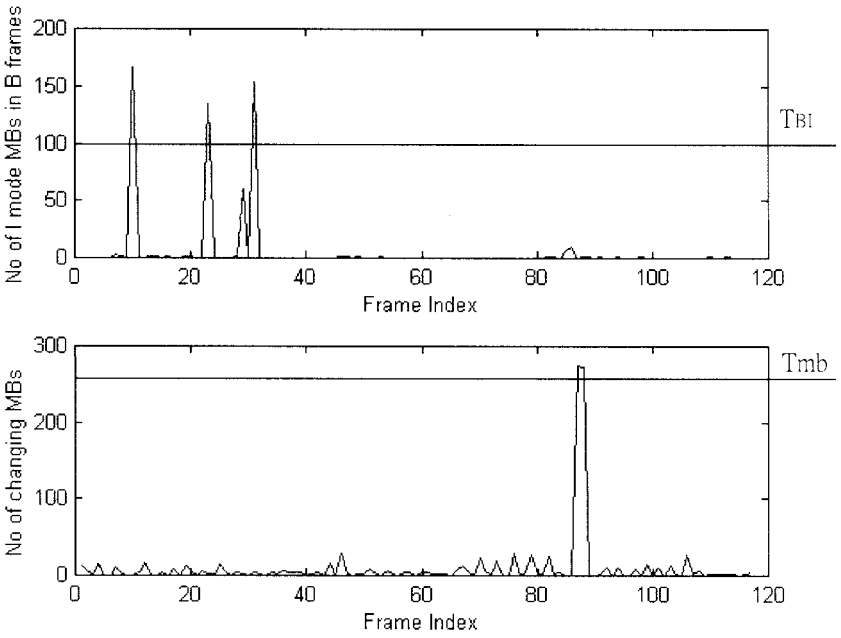


Fig. 10. Flashlight detection using news experiment sequence.

a significant number of MB's in the P_r frame are I mode similar to the case in the gradual scene change. As a result, the SGOP's in both fast panning and gradual scene change sequences are detected as potential SGOP's. The detection process may confuse fast panning operation with dissolve operation because of the similarity. Fortunately, the pattern of MB type in the *Complete Fast Panning SGOP* during the fast panning sequence is very different from that in the complete dissolve SGOP during the gradual scene change sequence. This difference can be used to tell these two sequences apart. In the complete fast panning SGOP, of which all four frames are involved in the panning sequence, B_f and B_r are more similar to P_f and P_r respectively under the assumption that panning is linear in time (frames). Therefore, more F mode MB's in B_f and B mode MB's in B_r is found. The patterns of MB types of a fast panning sequence are illustrated in Fig. 5.

C. Pattern of MB Types in SGOP's with Flashlights

The specific frame with flashlights would make itself very different from other frames in a SGOP. Fig. 9(b) shows an example of flashlights. If this specific frame appears at a B frame (B_f or B_r) in a SGOP, the large difference from both P_f and P_r in the SGOP makes MB's of this frames suitable for neither F nor B mode. In addition, flashlight would make the scene uniformly bright in some area. This property of uniformly brightness would cause this area of low variance. Thus, a significant number of MB's of this frame are inclined to be I mode which is rare in an ordinary video sequence. Another possibility is that the frame with flashlight lies at a P frame. In this case, abrupt scene change detection will find consecutive two scene changes in consecutive two SGOP's because two B frames previous to this specific frame will be inclined to F mode and two B frames following it will be inclined to B mode. This situation is unusual and can be used for extracting flashlights. The patterns of MB types of SGOP's with flashlights are illustrated in Fig. 6.

III. Experimental Result

A. Threshold Setting

We have edited a video sequence digitized from NTSC analog video signals of news, advertisement and sports programs. There are 24 abrupt scene changes and five gradual scene change transitions in this test sequence. In addition, there are also four shots involving camera motion (two fast panning sequences and two zooming sequences) and one shot with a large moving object occupying almost one-fourth of the frame. Each frame of the sequence is in SIF (352 $\times$ 240) format, which yields 330 (22 $\times$ 15) MB's in a frame. The frame structure of the test sequence is IBBPBBPBBPBBPBB ($M = 3, N = 15$) and the total bit rate is 1.5 Mbps. The experiments are performed using the encoder model of *Berkeley MPEG tools* (TM5, Vers. 1.0, Rel. 2; Aug. 1995) developed by L. A. Rowe *et al.* [16]. From these test sequences, we want to define and test a set of thresholds to apply to a much longer real video sequence described in Section III-B.

1) Abrupt Scene Change Detection: Fig. 7(a) shows the locations of detected changing MB's using the concept of MBGxy in the test sequence from Frame 1 to Frame 100, where each white dot represents a changing MB. In the abrupt scene change detection, the first step is to find the total number of changing MB's in each frame [shown in Fig. 7(b)]. Then, a user-defined threshold ($T_{\rm mb}$) is compared with the number and the one that is larger than $T_{\rm mb}$ is considered a scene change frame. Different $T_{\rm mb}$s are used and the experimental results are shown in Table I.

From Table I, it is found that $T_{\rm mb}$ can be chosen within the range from 270 to 300 without miss-detection or false

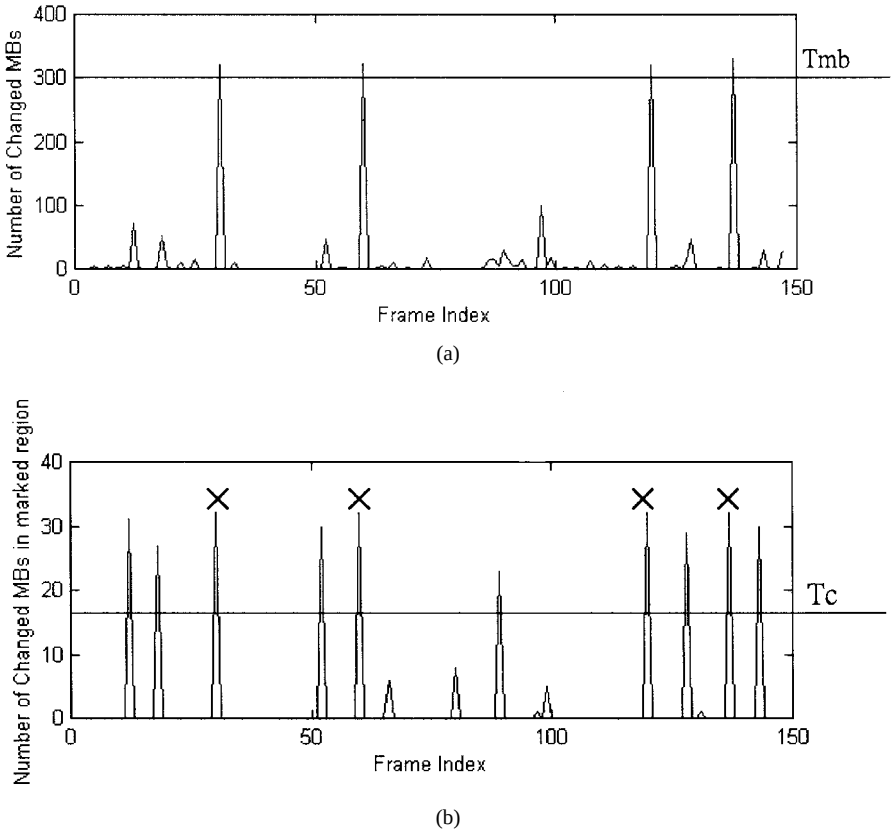


Fig. 11. Caption detection process. (a) Abrupt scene change detection. (b) Caption detection with eliminating the abrupt scene changes.

alarm. In fact, we found that $T_{\rm mb}$ appropriate for preventing miss detection and false alarm is from 257 to 301 on this test sequence. The wide range of $T_{\rm mb}$ indicates that the abrupt scene change detection is not strongly affected by the choice of $T_{\rm mb}$. The false alarms with small $T_{\rm mb}$ mainly come from starting or ending frames of fast panning sequences and gradual scene changes, or the complete fast panning SGOP's.

C. Gradual Scene Change Detection: As discussed in Section II-B, the P frame of SGOP's during gradual transition such as dissolve or fast panning sequence contains a significant number of I mode MB's. This characteristic can help us to detect the gradual scene change and fast panning sequence. However, it is worthy to note that the P frames of SGOP's containing abrupt scene changes also have very high ratio of I mode MB's. Therefore, the P frames with high ratio of I mode MB's resulting from abrupt scene changes have to be eliminated. The remaining SGOP's with significant number of I mode MB's in the P frame are defined as potential SGOP's. Threshold T_I is used to decide how many Intra-coded MB's in the P frame of a SGOP is significant to be a potential SGOP.

The second step is to distinguish each remaining potential SGOP coming from dissolve or fast panning. Utilizing the patterns illustrated in Fig. 4, the MB's in B frames of complete dissolve SGOP are inclined to be FB mode. As a result, a transition during several consecutive potential SGOP's are judged a dissolve sequence if the number of FB MB's in B frames exceeds a user-defined threshold T_{FB}. It is noted that a SGOP ending at I frame ($SGOP_2$) with adjacent potential SGOP(s) will also be regarded as a potential SGOP to prevent miss detection. The detection algorithm is illustrated in Fig. 8 and all five dissolve and two fast panning sequences are found.

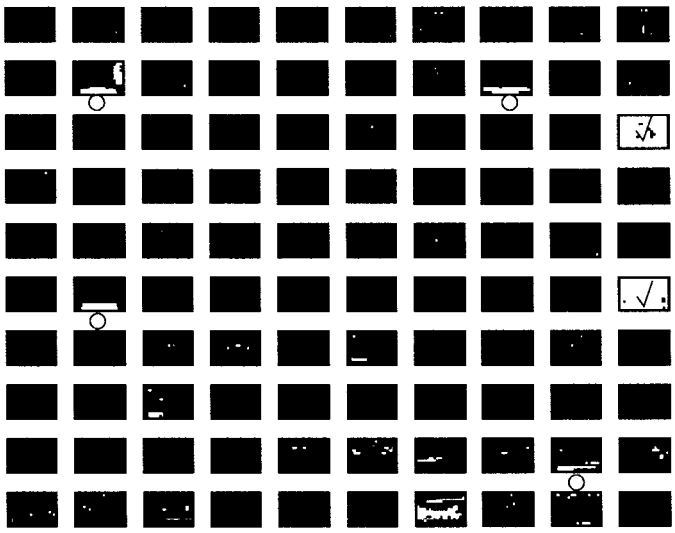


Fig. 12. The locations of changing MB's in the caption sequence (Frame 1 to Frame 100) and $\surd$ locates the scene change frame and ◯ locates the frame with captions.

An experiment of the influence of the threshold T_I and T_{FB} is performed and shown in Tables II and III. A range from 22 to 93 is appropriate choice of T_I and that from 220 to 320 is appropriate for T_{FB}. Similar to $T_{\rm mb}$, the detection algorithm also benefits from the wide range of choices of T_I and T_{FB}. The false alarms mostly result from large moving objects.

TABLE I
INFLUENCE OF T_{mb}

T_{mb}	Detected	Miss	False Alarm
330	10(41.7%)	14(58.3%)	0(0.0%)
315	21(87.5%)	3(12.5%)	0(0.0%)
300	24(100%)	0(0.0%)	0(0.0%)
285	24(100%)	0(0.0%)	0(0.0%)
270	24(100%)	0(0.0%)	0(0.0%)
255	24(100%)	0(0.0%)	1(4.2%)
240	24(100%)	0(0.0%)	3(12.5%)
165	24(100%)	0(0.0%)	6(25.0%)

TABLE II
INFLUENCE OF T_I

T_I	Dissolve		Fast Panning		False Alarm
	Detected	Miss	Detected	Miss	
10	5	0	2	0	4
20	5	0	2	0	1
30	5	0	2	0	0
50	5	0	2	0	0
80	5	0	2	0	0
100	4	1	2	0	0
165	3	2	0	2	0

TABLE III
INFLUENCE OF T_{FB}

T_{FB}	Dissolve	Fast Panning
320	5	2
310	5	2
300	5	2
250	5	2
220	5	2
200	6	1
165	6	1

TABLE IV
EXPERIMENTAL RESULTS USING SEQUENCES OF MOVIES

Type	**Seq 1**	**Seq 2**	**Seq 3**	**Seq 4**
Abrupt Scene change	66	78	75	50
Detected	66	78	75	50
False alarm	0	0	0	1
Miss	0	0	0	0
Gradual scene change	2	2	1	9
Detected	2	2	1	9
False alarm	0	1	1	3
Miss	0	0	0	0
Fast panning detection	4	6	3	19
Detected	4	5	2	21
False alarm	0	0	0	3
Miss	0	1	1	1
Caption detection	25	27	21	13
Detected	29	31	22	14
False alarm	4	5	4	2
Miss	0	1	3	1
Detected Flashlight	1	2	0	3

TABLE V
STATISTICS OF MBS IN B FRAMES

Type	***Attribute***			***Total***
FB mode	*Not Coded*	*Coded, Quant*	*Coded*	*Total*
1.5M	34942	26288	16115	**77345**
6.0M	5449	51128	20768	**77345**
F mode	*Not Coded*	*Coded, Quant*	*Coded*	*Total*
1.5M	5371	5054	6683	**17108**
6.0M	1692	12482	2934	**17108**
B mode	*Not Coded*	*Coded, Quant*	*Coded*	*Total*
1.5M	7787	2940	5911	**16638**
6.0M	1437	11258	3943	**16638**
I mode	*Not Quant*		*Quant*	*Total*
1.5M	120		329	**449**
6.0M	334		115	**449**

TABLE VI
STATISTICS OF MBS IN P FRAMES

Type	***1.5M***	***6.0M***
Intra	3023	7555
Intra,Quant	5785	1253
Total of I mode	**8808**	**8808**
No MC, Coded	8129	14169
No MC, Coded, Quant	9365	3439
Total of No MC	**17494**	**17608**
Fwd, Not Coded	540	354
Fwd, Coded, Quant	11038	2694
Fwd, Coded	7000	15416
Total of F mode	**18578**	**18464**

TABLE VII
STATISTICS OF MBS IN I FRAMES

Bit Rate	**Intra**	**Intra, Quant**	**Total**
1.5M	**5954**	**5266**	**11220**
6.0M	**10856**	**364**	**11220**

3) Flashlight Detection: There is only one frame containing flashlights in the experimental sequence. In order to obtain more reliable results, we adopt a news sequence with four frames containing flashlights extracted from news reports. In this four second-long sequence, President Lee of Taiwan walks out of the lobby. Among four flashlights, three lie in B frames and one in the P frame. From Fig. 10 the former three flashlights lying in B frames are detected because a significant number of MB's are found I mode. The last flashlight on the P frame makes consecutive two frames (frame87 and frame88) detected as abrupt scene change frames. It is appropriate from 53 to 126 for T_{BI}. When smaller T_{BI} is chosen, frame 29 is detected as a flashlight. It may not be regard as a false alarm because from Fig. 9(b), the flashlight in frame 29 is simply not as bright as flashlight frames (frame 31, for example). In other words, by adjusting the T_{BI}, different brightness of flashlights can be detected.

D. Caption Detection: The MBGxy in Fig. 3 can be used to detect the changing MB in the position (x,y). By applying this idea, the locations of changing MB's can be used for caption detection. We observe that captions most often appear at the bottom of the video frame [see Fig. 9(c)]. According to this observation, we can change our scene change detection from the entire frame to the bottom region of the frame.

TABLE VIII
DIFFERENT MB PATTERNS CORRESPONDING TO DIFFERENT SGOP STRUCTURE

Scene change detection: MB pattern modification corresponding to different GOP structures.
SGOP (P**BBP**): **Bold character means the scene change frame, flashlight frame or the gradual transition sequence.**
MB pattern : X: don't care, I,B,F,D: significant I,B,F,FB mode, **I,B,F,D**: dominant **I,B,F,FB** mode

SGOP(PBP)	P**BP**	PB**P**		
MB Pattern	X**B**I	X**F**I		
SGOP(PBBP)	P**BBP**	PB**BP**	PBB**P**	
MB pattern	X**BB**I	X**BF**I	X**FF**I	
SGOP(PBBBP)	P**BBBP**	PB**BBP**	PBB**BP**	PBBB**P**
MB Pattern	X**BBB**I	X**FBB**I	X**FFB**I	X**FFF**I
Complete dissolve sequence : MB pattern modification corresponding to different GOP structures				
SGOP(PBP)	P**BPBPBP**	B**PBPBPB**		
MB Pattern	X**D**I**D**I**D**I	XI**D**I**D**IX		
SGOP(PBBP)	P**BBPBBPBBP**	PB**BPBBPB**BP	PBB**PBBP**BBP	
MB pattern	X**DD**I**DD**I**DD**I	XF**D**I**DD**I**D**BI	XFFI**DD**IBBI	
SGOP(PBBBP)	P**BBBPBBBPBBBP**	PB**BBPBBBPBBB**P	PBB**BPBBBPB**BBP	PBBB**PBBBP**BBBP
MB Pattern	XX**D**XIX**D**XIX**D**XI	XF**DD**IX**D**XI**DD**BI	XFF**D**IX**D**XI**D**BBI	XFFFIX**D**XIBBBI
Complete fast panning sequence : MB pattern modification corresponding to different GOP structures				
SGOP(PBP)	P**BPBPBP**	B**PBPBPB**		
MB Pattern	XXIXIXI	FIXIXIB		
SGOP(PBBP)	P**BBPBBPBBP**	PB**BPBBPB**BP	PBB**PBBP**BBP	
MB pattern	XFBIFBIFBI	XFXIFBIXBI	XFFIFBIBBI	
SGOP(PBBBP)	P**BBBPBBBPBBB**P	PB**BBPBBBPBBB**P	PBB**BPBBBPB**BBP	PBBB**PBBBP**BBBP
MB Pattern	XFXBIFXBIFXBI	XFFBIFXBIFBBI	XFFXIFXBIXBBI	XFFFIFXBIBBBI
Flashlight detection: MB pattern modification corresponding to different GOP structures				
SGOP(PBP)	P**BP**	PB**P**BP		
MB Pattern	XIX	X**F**I**B**X		
SGOP(PBBP)	P**BB**P	PB**B**P	PBB**P**BBP	
MB pattern	XIXX	XXIX	X**FF**I**BB**X	
SGOP(PBBBP)	P**BBB**P	PB**BB**P	PBB**B**P	PBBB**P**BBBP
MB Pattern	XIXXX	XXIXX	XXXIX	X**FFF**I**BBB**X
Caption detection : The same as abrupt scene change detection				

However, the scene change frame has to be detected and skipped because it also contains many changing MB's in the bottom region. We have adopted a video sequence with four abrupt scene changes and six abruptly embedded captions for simulation. From Fig. 11, six captions are detected in marked region of {X-axis: 193 to 224, Y-axis: 49 to 304} after elimination of abrupt scene changes (The threshold is set as half the number of blocks in the marked region, $T_C = 16$). Fig. 12 shows the locations of changing MB's and gives a clearer illustration of the caption detection.

B. Experiments Using Real Video Sequence of Movies and News

In order to examine the application of proposed detection method, three video sequences extracted from movies and news were used. Each sequence is 9000 frames (5 min) long and each frame is 352 × 240 pixels in size. Table IV shows the experimental results using the threshold set of $\{T_{\rm mb} = 280, T_I = 40, T_{FB} = 300, T_{BI} = 100, T_C = 16\}$. It is found that all abrupt scene changes are successfully detected with only one false alarm. This false alarm comes from the sudden rapid explosion of the dynamite that creates large amount of smoke and dust covering most part of the frame. As to gradual scene change detection, some panning sequences with very homogeneous background become false alarms owing to their large number of FB mode MB's in B frames. This case also causes the miss detection of fast panning detection. In the caption detection, some smaller captions are miss-detected

TABLE IX
SIMULATION RESULT WITH RESPECT TO DIFFERENT SGOP STRUCTURE BASED ON THRESHOLD SET $\{T_{\rm mb} = 280, T_I = 40, T_{FB} = 300, T_{BI} = 100, T_C = 16\}$

SGOP	PBP	PBBP	PBBBP
Scene change	24	24	24
Detected	24	24	24
False alarm	0	0	0
Miss	0	0	0
Gradual scene change	4	4	4
Detected	4	4	4
False alarm	0	0	0
Miss	0	0	0
Fast panning	2	2	2
Detected	2	2	2
False alarm	0	0	0
Miss	0	0	0
Flashlight	4	4	4
Detected	4	4	0
False alarm	0	0	0
Miss	0	0	0
Caption	6	6	6
Detected	7	6	6
False alarm	1	0	0
Miss	0	0	0

TABLE X
COMPARISON OF DIFFERENT METHODS

Method	Extraction	Detection	Block	Accuracy	Application
DCT coefficients DC+AC	VLC, Inverse DPCM, Inverse Quantization	Operation(+,-) & Analysis	8×8	GOP	H.261 MPEG1,2
DCT coefficients DC only	VLC, Inverse DPCM	Operation(+,-) & Analysis	8×8	GOP	H.261 MPEG1,2
Motion Vectors	VLC, Inverse DPCM	Operation(+,-) & Analysis	16×16	SGOP	H.261 MPEG1,2
MB types	VLC	Counting & Analysis	16×16	Frame	MPEG1,2

and some moving objects occurring in the marked region cause false alarms. Several flashlights are existed in these sequences. Some of them come from photographic flashlights and others come from the sudden flashes of the light in the living room or the lamp at the desk.

IV. DISCUSSION

A. Influence of Bit Rates

Different bit rates may result in different occurrence frequency of MB types. Theoretically, more MB's are inclined to be I mode when coding at higher bit rates. In order to examine this assumption, an experiment of encoding the test sequence in different bit rates is done and the statistics of MB's is shown in Tables V–VII.

From Tables V–VII, it is found that there are only very slight changes in quantity of I, B, F, and FB types between two cases of bit rates. However, the difference occurs in the attribute of these four MB types. In the case of higher bit rate, MB's are inclined to be coded with default quantizer instead of being quantized using larger quantizer scale or not coded in order to improve the image quality. In other words, taking the Forward motion compensated MB for instance, the attribute of MB's change from "Not coded" or "Coded, Large Quantizer Scale" to "Coded, Default" to obtain better residue data reconstruction instead of changing FB to Intra mode. We also can find that due to increasing of the bit rate, MB's of I frames adopt the default quantizer. From the statistics, a conclusion for that the change of bit rates just very slightly affects the occurrences of I, B, F, FB modes is found. Therefore, the detection algorithm exploiting the MB types is not affected by bit rates.

B. Modification of Proposed Method Corresponding to Different GOP Structure

In the MPEG coding scheme, the GOP structure is not defined so that different GOP and SGOP structures can be adopted. For the encoding scheme containing only I and P frames, the proposed approach needs to be modified and integrated with other proposed method using DCT coefficients or motion vectors. Table VIII shows the modification of MB type patterns corresponding to different SGOP structures. Simulation results based on the same test sequence experimented in Section III-A, but with the different SGOP structures of Table VIII are shown in Table IX. From Table IX, we can conclude that our proposed method is practical for the coding scheme using bi-directional motion compensation and robust to different SGOP structures.

C. Advantage of Proposed Method

Experimental results have shown that the MB type information in the MPEG framework can be successful applied to achieve very fast scene change, flashlight and caption detection. This novel approach has some advantages worthy to be noted. First, the MB type information can be directly extracted from the bit streams of compressed video after VLC decoding. Second, detecting process is quite simple as compared to current DCT methods that require more operations. It needs to calculate the differences of DCT coefficients and accumulate these differences to create the interesting data to compare for scene change detection. However, only counting the MB types can provide us enough information for the same purpose. Third, its detection accuracy is on the frame scale, in other words, we can precisely indicate which frame the scene change occurs. In contrast, the method using DCT coefficients detects only one cut if there are two scene changes in a GOP. Last but not least, the proposed method has higher sensitivity that can detect the scene changes between two consecutive frames belonging to different shots with similar contents. The inclination of prediction utilized in our method is still valid because the almost identical frame would be predicted instead of the similar frame belonging to a different shot. In the other words, the macroblock information still follows some specific pattern when the abrupt scene change occurs. However, for current other detection methods, it makes setting the threshold either of DC difference or of motion value summation difficult on this occasion. It is noted that the proposed algorithm is useful only in the MPEG framework containing B frames. Integration with proposed algorithm using DCT coefficients is needed for the coding scheme containing only I and P frames. The comparison of several different detection methods is shown in Table X.

V. CONCLUSION

We have developed a novel video analysis method using MB type information. By this method, satisfactory detection

precision and speed is obtained. In addition, the method using MB type information benefits from easy data extraction from the bitstream, very simple analysis, frame-based accuracy and high sensitivity to avoid miss detection. For coding schemes containing only P or I frames, future work of integration of our method and other proposed methods is needed.

REFERENCES

[1] A. Nagasaka and Y. Tanaka, "Automatic video indexing and full motion search for object appearances," *Visual Database Systems, II*, E. Knuth and L.M. Wegner, Eds. Amsterdam, The Netherlands: North-Holland, pp. 119–133, 1991.

[2] F. Arman, R. Depommier, A. Hsu, and M-Y. Chiu, "Content-based browsing of video sequences," *ACM Multimedia'94*, pp. 97–103, Aug. 1994.

[3] F. Arman, A. Hsu, and M. Y. Chiu, "Feature management for large video databases," *Storage and Retrieval for Image and Video Databases*, vol. SPIE-1908, pp. 2–12, 1993.

[4] Y. Nakajuma, "A video browsing using fast scene cut detection for an efficient networked video database access," *IEICE Trans. Inf. Syst.*, vol. E77-D, no. 12, pp. 1355–1364, Dec. 1994.

[5] F. Arman, A. Hsu, and M. Y. Chiu, "Image processing on compressed data for large video databases," in *Proc. First ACM Int. Conf. Multimedia*, Aug. 1993, pp. 267–272.

[6] H. J. Zhang, A. Kankanhalli, and S. W. Smoliar, "Automatic partitioning of full-motion video," *Multimedia Syst.*, vol. 1, pp. 10–28, July 1993.

[7] H. J. Zhang, C. Y. Low, and S. W. Smoliar, "Video parsing and browsing using compressed data," *Multimedia Tools Applicat.*, vol. 1, no. 1, pp. 89–111, Mar. 1995.

[8] B.-L. Yeo and B. Liu, "Rapid scene analysis on compressed video," *IEEE Trans. Circuits Syst. Video Technol*, vol. 5, pp. 533–544, Dec. 1995.

[9] K. Otsuji and Y. Tonomura, "Projection detecting filter for video cut detection," in *Proc. First ACM Int. Conf. Multimedia*, Aug. 1993, pp. 251–257.

[10] S. F. Chang and D. G. Messerschmitt, "Manipulation and compositing of MC-DCT compressed video," *IEEE J. Select. Areas Commun.*, vol. 13, pp. 1–11, Jan. 1995.

[11] N. V. Patel and I. K. Sethi, "Compressed video processing for cut detection," *proc. Inst. Elect. Eng.—Vis., Image Singal Process.*, vol. 143, no. 5, pp. 315–323, Oct. 1996.

[12] N. V. Patel and I. K. Sethi, "Video shot detection and characterization for video databases," *Pattern Recognit.*, vol. 30, no. 4, pp. 583–592, Mar. 1997.

[13] B.-L. Yeo and B. Liu, "Visual content highlighting via automatic extraction of embedded captions on MPEG compressed video," *Proc. SPIE, Digital Video Compression: Algorithms Technolog.*, vol. 2668, pp. 58–47, 1996.

[14] J. Meng, Y. Juan, and S.-F. Chang, "Scene change detection in a MPEG compressed video sequence," *Proc. SPIE, Digital Video Compression: Algorithms Technol.*, vol. 2419, pp. 14–25, 1995.

[15] Q. Wei, H. Zhang, and Y. Zhong, "A robust approach to video segmentation using compressed data," *Proc. SPIE, Storage and Retrieval for Still Image and Video Database*, vol. 3022, pp. 448–456, 1997.

[16] L. A. Rowe, S. Smoot, K. Patel, B. Smith, K. Gong, E. Hung, D. Banks, S. T.-S. Fung, D. Brown, and D. Wallach, *Berkeley MPEG Tools*, Vers. 1.0, Rel. 2; Aug. 1995) [Online] Available ftp://mm-ftp.cs.berkeley.edu/pub/multimedia/mpeg/bmt1r1.tar.gz.

Soo-Chang Pei (S'71–M'86–SM'89) was born in Soo-Auo, Taiwan, R.O.C., in 1949. He received the B.S.E.E. degree from National Taiwan University (NTU), Taipei, in 1970 and the M.S.E.E. and Ph.D. degrees from the University of California, Santa Barbara, in 1972 and 1975, respectively.

He was an Engineering Officer in the Chinese Navy Shipyard from 1970 to 1971. From 1971 to 1975, he was a Research Assistant at the University of California, Santa Barbara. He was Professor and Chairman in the Department of Electrical Engineering, Tatung Institute of Technology, Taiwan, from 1981 to 1983. Presently, he is the Professor of Department of Electrical Engineering, NTU. His research interests include digital signal processing, image processing, optical information processing, and laser holography.

Dr. Pei is member of Eta Kappa Nu and the Optical Society of America.

Yu-Zuong Chou was born in Tainan, Taiwan, R.O.C. He received the B.S. degree from the National Tsing Hwa University, Hsinchu, Taiwan, in 1994, and the M.S. degree from the National Taiwan University (NTU), Taipei, in 1996, both in electrical engineering. He is currently pursuing the Ph.D. degree in electrical engineering at NTU.

He served in the Army as a Communication Officer from 1996 to 1998. His current research interests include video compression, image processing and multimedia application.