

## 多媒體訊號處理 (II) – 總計畫

### Multimedia Signal Processing (II)

計畫編號: NSC89-2213-E-002-159

執行期限: 89 年 8 月 1 日至 90 年 7 月 31 日

主持人 貝蘇章 台灣大學電機系教授

#### 摘要

隨著影像資料的漸趨龐大，一個有效的影像分析方法變得日益重要。而現階段，由於大部分的影像資料都以 MPEG 的方式壓縮儲存，所以一個直接在壓縮資料做的影像分析也引起廣泛的研究興趣。本文提供一個新的快速影像分析方法，以直接萃取影像資料中的 Macroblock 資料加以簡單分析，可以快速找到換景、閃光燈及標題等較關鍵的影像片段，以便利將來使用者的瀏覽與搜尋。

**關鍵字:** MPEG、影像分析、換景

#### 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.

**Keywords:** Compressed Domain Video Analysis, Video Indexing, Scene Change, Flashlight and Caption Detection.

#### 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 nowadays. Among 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 attracted a lot of research.

Up to now, many approaches have been developed to detect scene changes of the video sequences [1]-[8]. In order to lower the computation cost, more and 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 analyses. 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 are usually stored in 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 contents mathematically result in similar DCT coefficients. Arman et. al [3] have proposed a method to partition the motion JPEG video sequences and experimental results show that the relation of DCT coefficients can be used to detect scene changes. Concerning the MPEG compression 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 the 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. 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 as the boundaries of a shot usually needs more complicated analysis. Zhang et. al [7] have proposed a multi-pass, 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 as 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 MBs resulting from the interpolation characteristic of dissolve effect help us to distinguish gradual scene changes from fast panning sequences. Similarly, the frame containing flashlights would make itself different from others and also affect the MB type information. Significant number of intra-coded MBs coming from bright background of the flashlight can indicate the occurrence of the flashlight. Caption embedding would make the 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 indicates 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.

### 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.

### Reference

- [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, editors, 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, Aug. 1994, pp. 97-103
- [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 Dec 1994
- [5] F. Arman, A. Hsu, and M.Y. Chiu, "*Image processing on compressed data for large video databases*", 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 and Applications, 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 and Systems for Video Tech, vol. 5, NO. 6, Dec 1995, pp. 533-544

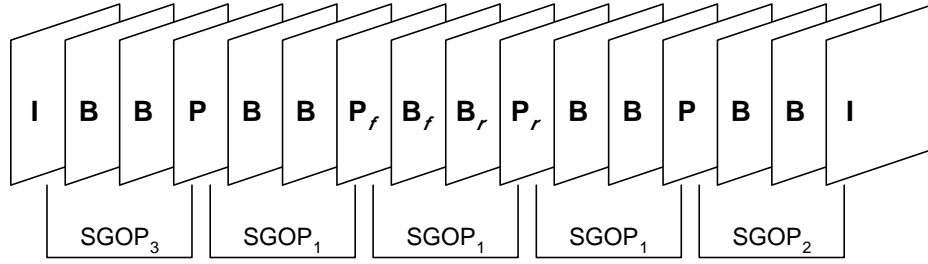


Fig. 1: An illustration of GOP structure used in the proposed method

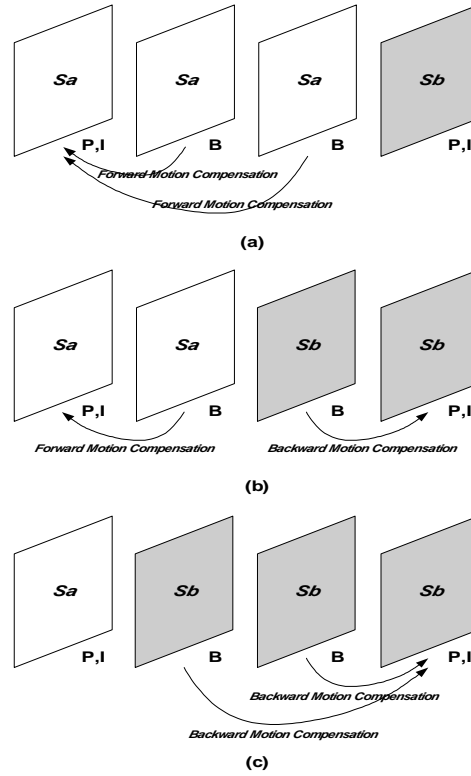


Fig. 2: Patterns of MB types in SGOPs with abrupt scene changes (a) Scene Change at P frame or I frame (SCPI) (b) Scene Change at Front B frame (SCFB) (c) Scene Change at Rear B frame (SCRB).

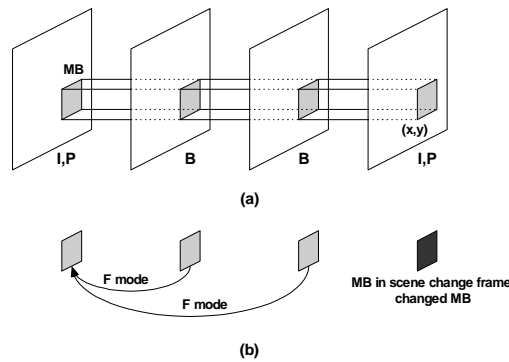


Fig. 3: (a) An illustration of MBGxy (b) An example of pattern of MB type in the case of SCPI



Fig. 4 (a): An example of gradual scene changes

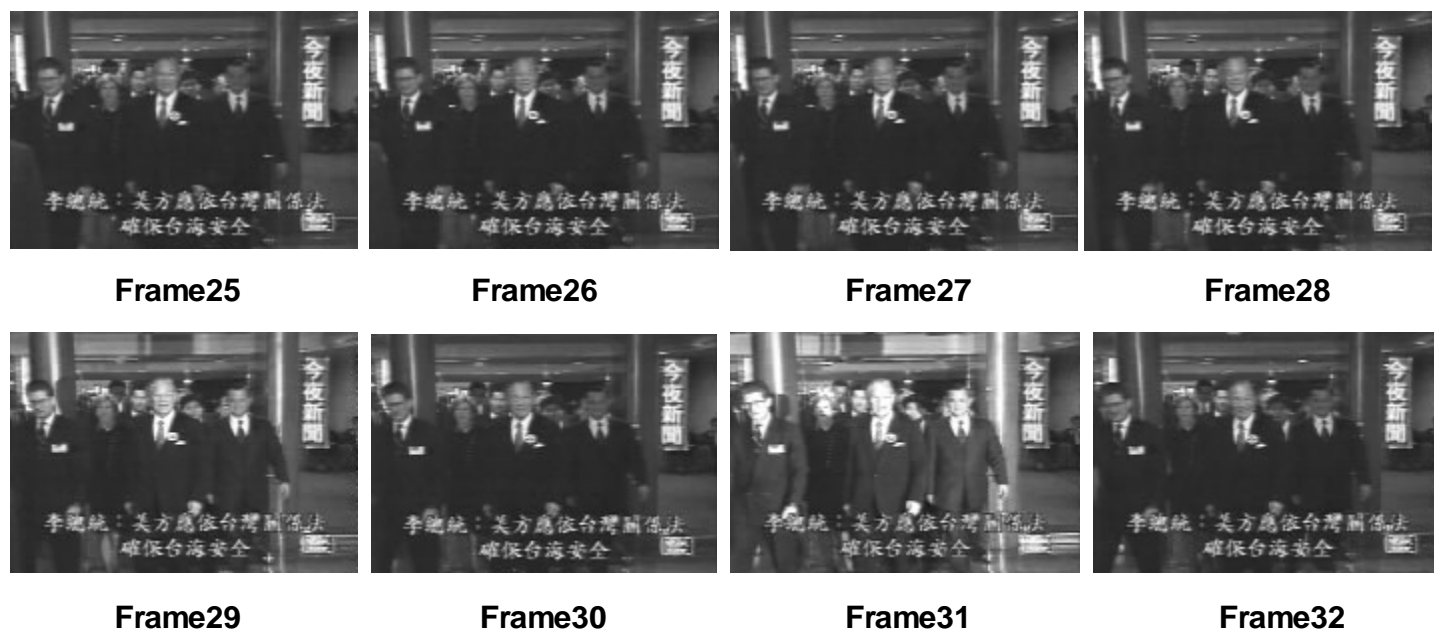


Fig 9(b): An example of flashlight in new experiment sequence (Frame 31)