SCENE CUT DETECTION FROM MPEG VIDEO STREAM CODED WITHOUT B PICTURES

Ali M Dawood

Etisalat College of Engineering
Emirates Telecom. Corporation
Sharjah, P.O.Box 980
United Arab Emirates
dawood@ece.ac.ae

Mohammad Ghanbari

Electronic Systems Engineering
University of Essex
Colchester CO4 3SQ
Essex - UK
ghan@essex.ac.uk

Keywords: Scene cut detection, Video data-base, video segmentation.

ABSTRACT

In this paper we propose an algorithm that automatically detects clear scene cut locations from an MPEG-1 video bit streams coded with a GOP structure of M=1, without B pictures. The algorithm detects scene cuts at P type pictures by monitoring the percentage of Intra-macroblocks per P picture. while scene cuts at I pictures are detected by matching the macroblocks type of the two P pictures at the GOP boundaries. A "Type Matching Parameter" (TMP) is developed to estimate the matching degree between the macroblock types of two P pictures. It is shown that the method is able to identify the location of scene cuts in P and I pictures with a high success rate.

1. INTRODUCTION

More and more video sequences are recorded and stored on storage media around the world at an incredible rate. This "treasure" of stored images requires special attention in order to maximise its value. In this context a collaborative effort has been spent on specifying the MPEG-7 standard, the ultimate solution for multimedia data-bases. One of the objectives set for MPEG-7 standard is to specify a standard set of descriptors that can be used to describe various types of multimedia information, that include high-level feature extraction [1]. Scene cut detection is a key issue in high-level video processing required for applications such as video browsing, video retrieval, video indexing etc.

Many scene cut detection methods proposed in the past were based on measuring the pixel intensity histogram of frame difference signals [2].

Although, these methods work well their usability is limited since they require MPEG video streams to be decoded first. It would make more sense if the scene cut could be detected from the information provided in the compressed bit stream.

On the compressed domain, a scene cut detection algorithm was proposed in [3] to detect clear scene cuts from MPEG-1 video bit streams coded with a "Group of Picture" (GOP) structure of M = 2, 3 and 4. These correspond to one, two and three B pictures, respectively, between every two anchor (I or P) pictures. The algorithm uses the percentage of Intra-coded macroblocks as scene cut indication and the motion vector types of B-frames to locate the scene cut in any frame. A similar method was proposed in [4], but using the percentages of interpolated, forward and backward predicted macroblocks in B pictures to detect scene cut from compressed MPEG video streams.

Although MPEG-1 supports I, P and B picture coding modes to achieve high compression ratio some applications may find it not suitable to employ all of them. As B pictures require re-ordering and out-of-sequence transmission of pictures [5] longer reconstruction delays and larger buffers at the receiver are inevitable, hence, they are not suitable for some real-time applications [6]. Since the algorithms proposed in [3] and [4] rely on the information in B pictures, they can not be employed to detect scene cuts from MPEG bit streams coded with a GOP structure of M = 1 due to the absence of B picture.

In this paper we propose a clear scene cut detection algorithm for MPEG-1 bit stream coded without B pictures. The elimination of B pictures from the GOP is achieved by setting the M parameter equal to 1. For example a GOP of N = 6 M = 1 will look like: $...P_6\ I_1\ P_2\ P_3\ P_4\ P_5\ P_6\ I_1\ P_2\ P_3\ ...etc$

2. SCENE CUT DETECTION ALGORITHM

In MPEG-1 video streams, coded with a GOP structure of M = 1, the scene cut in P pictures can be easily detected by monitoring the the percentage of Intra-coded macroblocks per picture. Any peak signal on this percentage is an indication of a scene cut [3]. However scene cuts in I pictures cannot be detected using the algorithm of [3] because the B pictures, which were extensively used to detect scene cuts in I pictures, do not exist. The method in [4] is also not suitable for the same reason. In this situation the different coding modes of P picture macroblocks are therefore the only source of information that need to be exploited in order to detect the scene cut in I pictures.

Since a video shot is defined as a continuous camera operation from start to stop, there may be zooming and panning but the variations are smooth [7], the characteristics of the pictures within a shot is expected to be similar to each other. In other words, the spatial difference between two adjacent pictures within a shot is small. Therefore, it is reasonable to argue that a coded macroblock in a picture n is more likely to be coded in picture $n + 1$ with the same type, provided that pictures n and $n + 1$ are of the same type and both fall within the same video shot. For example in a video shot, which is coded with a GOP parameter of M = 1, if the coder decided that a macroblock in a P_n picture is motion compensated but not coded (MC/NC) type then in the next P picture , P_{n+1}, that macroblock is more likely to be of MC/NC type, because both pictures have a similar spatial characteristics.

Based on this fact, in MPEG-1 stream coded with a GOP structure of N = 12 and M = 1, a scene cut in an I picture can be detected by matching the macroblocks types of the first P picture (P_2) in a GOP with their counterparts type of the last P picture (P_{12}) in the previous GOP, as illustrated in figure 1. Remember that the GOP start with an I picture (I_1). The matching results indicate whether both (P_2) and (P_{12}) fall into the same video shot or not.

At the beginning a macroblock (MB) type matrix is developed to collect the type of the MBs within a picture. Let $T^+ = \{t^+(i,j)\}$ and $T^- = \{t^-(i,j)\}$ be two MB type matrices for the P pictures directly after and before an I picture respectively, where $t^+(i,j)$ represents the $(i,j)^{th}$ MB type in a P picture immediately following an I picture, and $t^-(i,j)$ represents the $(i,j)^{th}$ MB type in a P picture immediately preceding the I picture. Then the MB type matching parameter (TMP) value of the two P pictures can be considered as the ratio between the number of matched MBs to the total number of MB within a picture. For a video picture of 352x288 pixels, with a total of 396 MB per picture, the TMP of T^+ and T^- is calculated by:

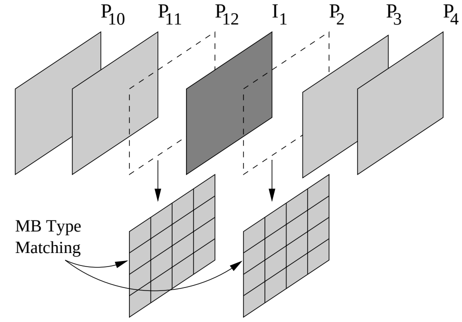


Figure 1: Macroblock type matching

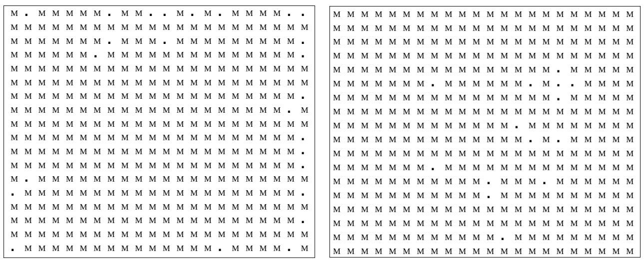
$$TMP(T^+, T^-) = \frac{\sum_{i,j} m\left(t^+(i,j), t^-(i,j)\right)}{396} \qquad (1)$$

where TMP can have any value between 0 and 1. The function $m(\;)$ is equal to binary "1" if $t^+(i,j)$ is identical to $t^-(i,j)$ and equal to "0" otherwise. Consequently, if most of the MBs in both pictures are identical then the function $m(\,)$ will give the result "1" for most of the MBs resulting in a high TMP value, hence it is an indication that the two P pictures belongs to the same video shot. Conversely, low TMP value will result from the case when each P picture belongs to a different scene.

For illustration, a test sequence was edited from two shots, for simplicity named as "Kids" and "News", as shown in figures 2 (a) and (b). The shot "Kids" focuses on three children sitting on a patterned sofa while watching TV, whereas "News" represents a normal News reader with a smooth background. Furthermore, the start of the "News" shot was located so that it is coded as an I picture when using an MPEG-1 encoder with a GOP structure of N = 12 and M = 1. In order to detect the scene cuts at I pictures the function $m\left(t^+(i,j), t^-(i,j)\right)$ was applied on every two P pictures appear before and after every I picture. The result of $m\left(t^+(i,j), t^-(i,j)\right)$ at an arbitrarily selected I picture from each shot is depicted in figures 3 (a) and (b), where the character "M" represents a match between two MB types while the "." (dot) denotes MB type mismatch. Figures 3 (a) and (b), have recorded a TMP value of 0.93 and 0.96 respectively. This is proves that most of the MBs of two P pictures within the same shot are coded with the same MB type. However, in the case of a scene cut between the "Kids" and "News" shots, the TMP value was 0.15 since most of the MB in the two P pictures are not of the same type, as shown in figure 4,

(a) Shot "Kids" (b) Shot "News"

Figure 2: Sample image from two different video shots



(a) Shot "Kids" (b) Shot "News"

Figure 3: MB type matching between two P pictures in the same shot

Therefore, calculating the TMP around all the I pictures will indicate whether the two P pictures (before and after the I picture) are in the same video shot or not, consequently deciding whether the I picture is a scene cut picture or not. However, if one of these P pictures are actually a scene cut picture then the TMP will be low indicating a scene cut in I picture. To remove this confusion, since scene cut in P pictures can be easily detected by monitoring the percentages of Intra-coded MB type, then a precise decision can be achieved when combining the TMP measurement with the Intra-coded MB percentages. In other words, if a majority of Intra-coded MBs are detected on one of the P pictures immediately before or after an I picture, then the process of TMP measurement is not performed around that particular I picture with the assumption that a scene cut already exist in one of the P pictures.

3. SIMULATION RESULTS

To evaluate the performance of the proposed scene cut detection algorithm, three test sequences were encoded using MPEG-1 encoder with a GOP structure of N = 12 and M = 1. The sequences were edited so that some of the I pictures coincide with the scene cut picture. Figure 5 depicts a snapshot from the "Test 1" sequence with 3 scene cuts between four video shots. Note that the first scene cut was at picture number 84 and it was

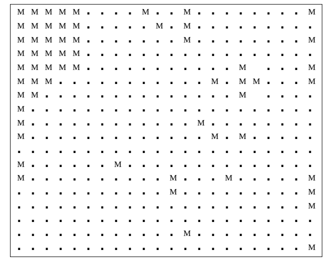


Figure 4: MB type matching between two P picture, one from shot "Kids" and the other is from shot "News"

an I picture while the other two were at picture number 172 and 257 and both were of P type.

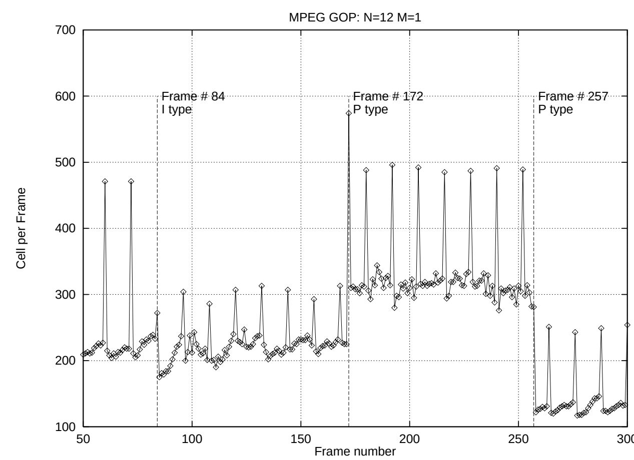


Figure 5: Frame 50-300 of the MPEG coded sequence

The algorithm is applied in two phases; firstly detecting scene cuts at P pictures by monitoring the majority of Intra-coded MB in every P picture. Secondly, calculating the TMP between the two P pictures surrounding every I picture. For the latter phase, it can be seen in figure 6 that a sharp drop in the TMP trace can be detected at the 7^{th} I picture indicating a scene cut at I picture number 84 in the original sequence.

The Intra-coded MB threshold for this particular test sequence was selected to be = 80%, while the TMP threshold was set to 0.6 . These two parameters was adjusted in the other two sequences for the optimum results. Table 1 and 2 list these two parameters setting along with the simulation results.

As it can be seen from table 1, the scene cuts at P pictures were correctly detected with 100%, 90% and 87% for "Test 1", "Test 2" and "Test 3" sequences respectively. The false scene cuts detected in "Test 2" and "Test 3" sequences were due to the high activity of their contents. This can be relieved by increasing the percentage of Intra-MB threshold, however genuine scene cut

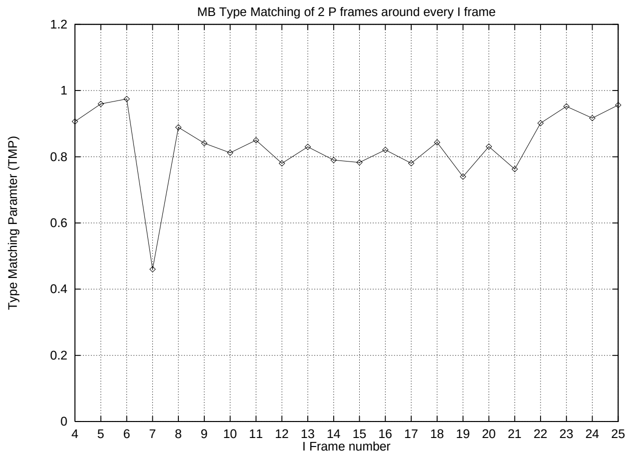


Figure 6: Type Matching Parameter (TMP) Trace

Sequence Name	*Int-MB Thresh.*	*Actual S. cut*	*Correctly Detected*	*False Detection*
Test 1	70%	20	20	0
Test 2	80%	22	22	2
Test 3	70%	30	27	1

Table 1: Scene cut detection at P pictures

may not be detected when the Intra-MB threshold is set to a very high value. In table 2, scene cuts at I pictures were correctly detected with 80% 77% and 76% for "Test 1", "Test 2" and "Test 3" sequences respectively. The detection accuracy is lower in "Test 3" sequence because of the higher activity of its content compared with the other two test sequences. The missed scene cuts at I pictures were due to the similarity in the global characteristics between the two shots, causing the MPEG encoder to produce similar MB type in the P pictures although they fall into two different scenes. The detection accuracy calculated above was based on deducting the the number of false detection from the number of correctly detected scene cuts.

4. CONCLUSIONS

We have proposed a scene cut detection algorithm that can be applied directly to the MPEG bit stream without the need to decode the frames. The algorithm was dedicated for MPEG streams that has been coded with a GOP structure of M = 1. The scene cuts at P pictures are detected by monitoring the percentage the Intra coded macroblocks, while a TMP (Type Matching Parameter) was developed to detect scene cuts at I pictures. The TMP is used to evaluate the degree of similarity between two P pictures, before and after an I picture. High TMP value indicates a high degree of similarity and hence no scene cut at I picture, and vice versa.

Sequence Name	*TMP Thresh.*	*Actual S. cut*	*Correctly Detected*	*False Detection*
Test 1	0.6	20	16	0
Test 2	0.8	18	15	1
Test 3	0.7	25	21	2

Table 2: Scene cut detection at I pictures

Almost all scene cuts at P picture were detected, while 76%-80% of the scene cuts at I picture were detected. The detection accuracy at I picture was not very high due to the similarity in the global characteristics between the successive shots, causing the MPEG encoder to produce similar MB type in the P pictures although they fall into two scenes. An optimisation is required between the two parameters used, i.e. the Intra-MB percentage and the TMP threshold.

5. REFERENCES

[1] ISO/IEC, "MPEG-7 context and objective," tech. rep., Coding of Moving Pictures and Audio, Version 6 San Jose, Feb 1998. ISO/IEC JTC1/SC29/WG11 N2082.

[2] S. Smoliar and H. J. Zhang, "Content-Based Video Indexing and Retrieval," *IEEE Multimedia Magazine*, vol. 1, pp. 62–72, Summer 1994.

[3] A. Dawood and M. Ghanbari, "Clear scene cut detection directly from mpeg bit streams," in *Seventh International Conference on Image Processing and its Applications, IPA'99*, (Manchester, UK), IEE, July 1999.

[4] W. Fernando and C. Canagarajah and D. Bull, "Sudden scene change detection in MPEG-2 video sequences," in *IEEE International workshop on multimedia signal processing, MMSP'99*, Copenhagen - Denmark, September 1999.

[5] ISO/IEC, "Generic coding and moving pictures and associated audio." Recommendation H.262, March 1994.

[6] P. Pancha and M. E. Zarki, "MPEG coding for variable bit rate video transmission," *IEEE Communications Mgazine*, May 1994.

[7] M. M. Yeung and B. Yeo, "Video visualisation for compact presentation and fast browsing of pictorial content," *IEEE Trans on circuits and sys for video tech*, vol. 7, pp. 771–785, oct 1997.