International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:11, 2008

Shot Transition Detection with Minimal Decoding of MPEG Video Streams

Mona A. Fouad, Fatma M. Bayoumi, Hoda M. Onsi, and Mohamed G. Darwish

***Abstract*—**Digital libraries become more and more necessary in order to support users with powerful and easy-to-use tools for searching, browsing and retrieving media information. The starting point for these tasks is the segmentation of video content into shots.

To segment MPEG video streams into shots, a fully automatic procedure to detect both abrupt and gradual transitions (dissolve and fade-groups) with minimal decoding in real time is developed in this study. Each was explored through two phases: macro-block type's analysis in B-frames, and on-demand intensity information analysis.

The experimental results show remarkable performance in detecting gradual transitions of some kinds of input data and comparable results of the rest of the examined video streams. Almost all abrupt transitions could be detected with very few false positive alarms.

***Keywords*—**Adaptive threshold, abrupt transitions, gradual transitions, MPEG video streams.

I. Introduction And Related Work

THE extensive amount of media coverage today, generates difficulties in identifying and selecting desired information. Digital libraries become more and more necessary in order to support users with powerful and easy-to-use tools for searching, browsing and retrieving of information. The starting point for these tasks is the segmentation of video content into shots.

To segment MPEG video streams into shots, a fully automatic procedure to detect both abrupt and gradual transitions (dissolves and fade-groups) with minimal decoding in real time is developed in this study. The detection of shots is explored by detecting transitions separating them, which are to be hard or gradual.

The abrupt transition detection is explored by first examining the number of forward and backward macro-blocks (p- and b-MBs) in consecutive B-frames. Then an intensity histogram comparison is applied in a novel appropriate way to confirm detected transitions.

The gradual transition is detected by first examining the intra-coded predicted macro-blocks (i-MBs) within successive B-frames. Then the detection is confirmed by checking the parabolic shape of the frame variants of the candidate sequence.

Several reviews and evaluation work of shot transition detections in uncompressed and compressed domains are available [1] [2] [3].

Works related to the presented study is found in [4], [5] [6] and [7]. Motion information in the MPEG video streams is the basic of their detection methodologies. In [4], the number of bi-directional predicted macro-blocks (bi-MBs) within a B-frame and their mean are used to define an adaptive threshold to identify abrupt scene changes. Dissolve and fade transitions are detected by dividing the B-frames into non-overlapping square areas, then the spatial distribution of bi-MBs in these sub-images are analyze, assuming that when a sequence has a gradual transition, the number of bi-MBs in each sub-image is similar with one another. J. Calic [5] introduces a method based only on temporal distribution of macro-block types. Motion vectors magnitude and directions are used to detect gradual transitions in [6]. In [7] a method for shot boundary detection of discrete cosine transform (DCT)-based and low-bit-rate encoded clips, which exploits the perceptual blockiness effect detection on each frame without using any threshold parameter is introduced.

The objective of this paper is to utilize available information in the compressed video streams to detect both abrupt and gradual transitions. In this work, MPEG-1 and MPEG-2 are to be considered (defined in ISO/IEC International Standards 11172 and 13818, respectively).

In comparison to those mentioned approaches, the results of the developed techniques are comparable, achieving the best balance between precision and recall for both abrupt and gradual transitions detection.

II. MPEG Overview

The compressed video data contains useful information that could be extracted easily without full-frame decompression, otherwise only minimum decoding is needed. The use of this information saves both time to perform full decompression and the additional storage for holding decompressed data.

The syntax of the MPEG standard video defines three main types of coded pictures: I-, P-, and B-frames, organized into Groups Of Pictures (GOP) sequence. Each GOP starts with an anchor frame (I- or P-frame) followed by two B-frames. An I-Frame is completely intra-coded. A P-frame is predicatively coded with reference to a past anchor frame, which is an I-, or P-frame. A B-frame is predicatively coded with bi-directional reference to a past and a future anchor frame.

Each frame is divided into non-overlapping 16x16-pixels macro-blocks (MBs) that. Each macro-block contains information about the type of its temporal prediction for motion compensation stored in its motion vector. There are four macro-block types; intra-coded (i-MBs), forward predicted (p-MBs), backward predicted (b-MBs), and bi-directional interpolated (bi-MBs).

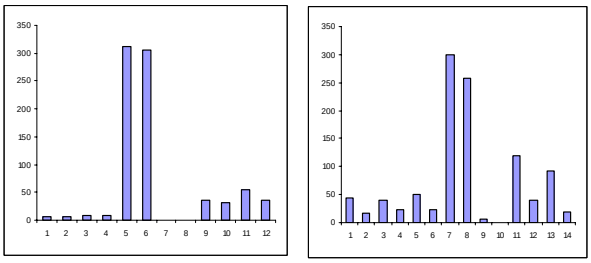
Macro-blocks of the predicted frames are matched to those in the anchor frame(s) for best match. If an identical MB exists in the typical position of an anchor frame, the MB is totally skipped. If the MB is similar with a MB in the anchor frame, the displacement between the positions of the two MBs is stored in its corresponding motion vector. If there is no similarity with any MBs in the anchor frame(s) at any position, the MB is then intra-coded [8].

III. Shot Transition Detection

In the presented work, two strategies are developed to detect abrupt and gradual transitions with minimal decoding of the MPEG video streams without any prior-information or fixed thresholds.

A. The Abrupt Transition Detection

This process is developed through two phases. First, a pre-detection phase is explored to select candidate frames by examining the number of forward and backward macro-blocks (P-, and b-MBs) in consecutive B-frames. Second, the intensity histogram differences of the candidate frames are examined in a novel way to confirm detected transitions. The developed abrupt transition detection strategy combining forward and backward MBs-types analyses with the intensity histogram comparison is very efficient working in real time.



(a) Number of b-MBs (b) Number of p-MBs

Fig. 1 Forward and backward MBs in B-frames

Phase-1: Analysis of Macro-Block Types

If an abrupt transition occurs at a B-frame, the number of the backward predicted macro-blocks (b-MBs) is significantly increased (Fig. 1-a). If the abrupt change occurs at an I- or P-frame, the preceding two B-frames record a significant increase in the number of the (p-MBs) forward predicted macro-blocks (Fig. 1-b). As shown in the figure, the highest values in each chart indicate the positions of the abrupt changes.

Practically, the number of the forward MBs is compared to a threshold Tp, while the number of backward MBs is compared to a threshold Tb. Both thresholds are adaptively estimated using the sliding-window technique [6].

Phase-2: Analysis of Frame Intensity Histogram

The detection of abrupt changes based on the analysis of macro-block types (phase-1) is not sufficient generating a lot of false positives. It should be supported by a more precise method that determine accurately the position of the abrupt change and prevent the appearance of false positive alarms. Hence, phase-2 is proposed to finally decide whether the candidate frame is really a transition or not. The idea behind performing the detection in two phases is to save time estimating histogram difference between each two consecutive frames. Otherwise it is estimated only on demand; when there is a considerable change of a frame that is recorded through changes of the number of predicted macro-blocks.

A main source of false positive alarms is the flashlights, which is defined as a sudden change of illumination in a single or few consecutive frames. A bi-directional intensity histogram analysis is performed to confirm the detected abrupt transitions from the first phase getting over flashlight effects. In the developed bi-directional procedure, the difference of the candidate frame and its preceding frame is compared to two thresholds T_{hb} and T_{ha} estimated using two n-frames temporal windows; before and after the candidate frame. The thresholds are adaptively estimated using the sliding window technique.

The Bi-Directional Procedure

1. Estimate the intensity histogram difference (using absolute difference) between the candidate frame and its adjacent preceding frame.
2. Compare the estimated difference to the thresholds T_{hb} and T_{ha}. If it exceeds them, an abrupt transition is confirmed.

The Adaptive Threshold Estimation

The adaptive threshold estimation for the histogram analysis begins by estimating the histogram difference between each two consecutive frames in an n-size temporal window. The value of n should be between one and two GOP-size, assuming that a shot duration is not less than 2-GOPs (n is chosen to be 15).

$$T_h = \mu + \alpha\sigma \tag{1}$$

Where μ and σ are the mean and standard deviation of the histogram differences in the temporal window, respectively, and α is a constant that should be selected carefully ($\alpha = 5$, according to Zhang's suggestion [9]).

The threshold T_h is called T_{hb} for the window before the candidate frame, and T_{ha} for a window after it.

With respect to the analysis of macro-block types, two thresholds are needed:

$$T_b = \sum_{i=1}^{c} \frac{NB_i}{(NB_i + NP_i)} \tag{2}$$

$$T_p = \sum_{i=1}^{c} \frac{NP_i}{(NB_i + NP_i)} \tag{3}$$

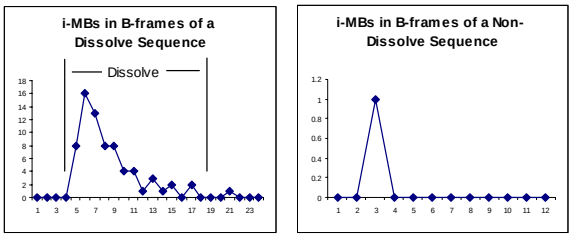
NP and *NB* are the number of forward and backward macro-blocks, respectively of B-frames in the n-frames temporal window having a total of *c* B-frames.

IV. Gradual Transition Detection

The gradual transition is detected through two phases. In the first phase the presence of intra-coded macro-blocks (i-MBs) within successive B-frames is examined. Then the detection is confirmed in the next phase by checking the parabolic shape of the frame variance of the candidate sequence. The detection and confirmation phases are applied in this study in a novel real time way. The novelty lies in: 1) there is no thresholds used at all, 2) the analysis of intra-coded MBs in B-frame to detected gradual sequences is explored for the first time, 3) the frame variance check is applied in real time.

Phase - 1: Analysis of intra-coded MBs

Frames in a single shot are almost identical unless there is a significant change due to camera or object movements or considerable change in lighting condition. On the other hand, the gradual transition sequence results from intensity, spatial and combined edits.



(a) A dissolve sequence (b) A non-dissolve sequence

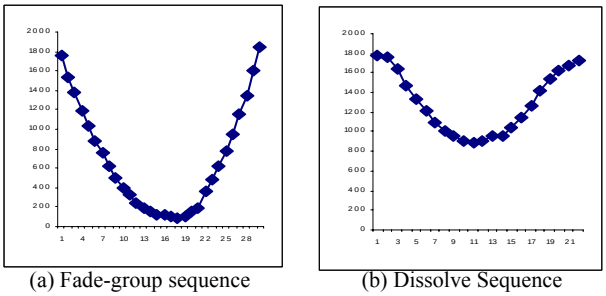
Fig. 2 i-MBs in consecutive B-frames

Since the B-frames have high temporal redundancy within a shot, a continuously strong motion predicted reference (b-, p-, or bi-MBs) will be present in the sequence as long as no significant changes occur in the scene. The intra-coded macro-blocks (i-MBs) begin to appear continuously in the B-frames once the significant changes present in the sequence (Fig. 2).

The developed detection strategy starts by searching for the beginning appear of i-MBs in the successive B-frames. If they continuously presents in the consecutive B-frame, the sequence is selected to be candidate for phase-2.

Phase - 2: Analysis of frame variance

Fig. 3 shows a typical parabolic shape of the frame variances of dissolve sequence and a variance curve for sequence with fade-out followed by an immediate fade-in (fade group) which is similar to the dissolve case except the valley goes down deeper to almost zero.



(a) Fade-group sequence (b) Dissolve Sequence

Fig. 3 A typical frame variance

The frame variance of the candidate sequence provided by phase (1) is examined to be of a parabolic shape in almost real time. The examination is applied by checking the increase and then decrease of the frame variance values within the sequence, having the smallest value nearly at the middle.

V. Results

To examine the developed shot detection technique, three sets composed of 12 natural TV video streams of different kinds (news, soccer, music, and movies) in each set, are used. The three sets have 88,454 total numbers of frames and image size of 352x240-pixels at frame rate of 25 fps. All sets have overall 739 abrupt transitions and 117 gradual transitions, most of them are of dissolve type and few fades.

All sets have several varieties in illumination, object and camera motion as well as the presence of flashlights. Although all video streams are of real source, they vary according to the extreme variations they have.

The testing data are grouped into three sets. Both sets (1) and (2) have abrupt transitions only. Set (3) have both kinds of transitions. Set (2) and set (3) have a lot of difficulties: 1) the extreme of changing in lighting conditions, 2) the blurring effect along with fast camera motion, 3) flashlights along several consecutive frames.

Analyzing the presented results, the few missed abrupt transitions are presented due to similar intensity distribution of frames in successive shots, which is very rare to be in a stream. On the other hand, false positives are due to presence of flash lights in several successive frames. However, the overall performance among the three sets is very promising.

Gradual transition detection is difficult to be detected due to object, camera operations and the blurring effects. The developed technique overcomes most of these difficulties providing a comparable performance through the two proposed phases.

The following two tables contain the summaries of the experimental results. Tables I and II contain the results of the developed abrupt transition detection procedure that applied to sets (1) and (2). Results of applying the developed gradual transition procedure to set (3) are shown in Table II.

TABLE I
THE PERFORMANCE OF THE DEVELOPED ABRUPT TRANSITION DETECTION PROCEDURE

Movie	Frames	Hard Cuts	Detected	Missed	False Positive
1	4095	27	27	0	0
2	4298	32	32	0	2
3	5493	45	45	0	3
4	6720	104	98	6	16
5	7968	47	44	3	0
6	6972	51	47	4	5
	35546	306	293	13	26
Precision =	0.9185	Recall =	0.95752		
7	6768	36	31	5	6
8	3625	57	50	7	6
9	5617	85	63	22	1
10	5554	107	83	24	11
11	8268	147	104	43	3
	29832	432	331	101	27
Precision =	0.9246	Recall =	0.7662		
Over all	65378	738	624	114	53
Precision =	0.9217	Recall =	0.84553		

TABLE II
THE PERFORMANCE OF THE DEVELOPED GRADUAL TRANSITION DETECTION PROCEDURE

Movie	Frames	Gradual	Detected	Missed	False Positive	Precision	Recall
*5	7968	16	15	1	5	0.7500	0.9375
*6	6972	48	40	8	37	0.5195	0.8333
12	8136	53	50	3	117	0.2994	0.9434
Over all	23076	117	105	12	159	0.3977	0.8974

Although the work presented in [6] used camera operations detection to reduce false alarms, the presented work achieved better recall and precision [10]. The recall evaluates how much the system could correctly detect shot transitions. While precision evaluates the produced false alarms. Table III shows the comparison.

The most important comparative result is that more than 90% of the frames in the gradual sequence could be detected by the proposed gradual transition detection procedure. It is seen from the table that the achieved recall for the gradual transitions detection is high due to following the developed intra-coded macro-block analysis of the successive B-frames.

TABLE III
THE COMPARATIVE RESULTS

	Abrupt Change		Gradual Change	
	Recall	Precision	Recall	Precision
Proposed	84%	92%	90%	40%
Doulaverakis [6]	80%	80%	38%	87%

VI. CONCLUSION AND FUTURE WORK

Concluding the results, the novel strategy to detect and confirm abrupt transitions provide a promising performance. The detection process is applied first with minimum decoding searching for significant change of the number of MB-types in B-frames. When it is found, the developed bi-directional histogram difference analysis is explored. It is very efficient eliminating false positives due to flashlights.

Furthermore, the detection of gradual transitions is also novel based mainly on analyzing the presence of intra-coded macro blocks in B-frames. The candidate sequences are confirmed using a real time check for the frame variance.

The proposed procedures were tested on numerous sequences containing a large variety of different kinds of video streams. The results are comparable to current literature.

However, future work may be conducted using additional MPEG features which should make the proposed procedures less sensitive to strong motion during shot changes without losing the real time capabilities.

REFERENCES

[1] R. Lienhart, "Reliable Transition Detection in Videos: A Survey and Practitioner's Guide," International Journal of Image and Graphics, 1(3):469-486, July 2001.
[2] I. Koprinska and S. Carrato, "Temporal video segmentation: a survey," Signal Processing: Image Communication 16(5):477–500, January 2001.
[3] U. Gargi, R. R. Kasturi, and S. H. Strayer, "Performance Characterization of Video-Shot-Change Detection Methods," IEEE Transactions on Circuits and Systems for Video Technology, vol. 10, No. 1, February 2000.
[4] S. Lee and M. H. Hayes, "Scene Change Detection Using Adaptive Threshold and Sub-Macro-block Images in Compressed Sequences," IEEE International Conference on Multimedia and Expo, 2001.
[5] J. Calic and E. Izquierdo, "Towards Real-Time Shot Detection in the Mpeg-Compressed Domain," Proceedings of WIAMIS'2001 - Workshop on Image Analysis for Multimedia Interactive Services, Tempera, Finland, May 2001.
[6] C. Doulaverakis, S. Vagionitis, Michalis Zervakis, and Euripides G. M. Petrakis, "Adaptive Methods for Motion Characterization and Segmentation of MPEG Compressed Frame Sequences," ICIAR (1) 2004: 310-317
[7] H. Koumaras,a G. Gardikis,b G. Xilouris,a E. Pallis,c and A. Kourtis, "Shot boundary detection without threshold parameters," 2006 SPIE and IS&T.
[8] D. LeGall, J. L Mitchell, W. B. Pennbaker, and C. E. Fogg, "MPEG video compression standard", Chapman & Hall, New York, USA, 1996.
[9] H.J. Zhang, A. Kankanhali, and S.W. Smoliar, "Automatic Partitioning of Full Motion Video. Multimedia Systems," 1993.
[10] S. Lefevre, J. Holler and N. Vincent, "Real-time Temporal Segmentation of Compressed and Uncompressed Dynamic Color image Sequences," International Workshop on Real Time Image Sequence Analysis, pages 56-62, Oulu, Finland, August 2000.