International Journal of Pattern Recognition
and Artificial Intelligence
Vol. 17, No. 8 (2003) 1467–1481

A NEW SLOW-MOTION REPLAY EXTRACTOR FOR SOCCER GAME VIDEOS

EN-JUNG FARN*, LING-HWEI CHEN† and JENN-HANN LIOU‡

Department of Computer Science and Information Engineering,
†Department of Computer and Information Science,
National Chiao Tung University, 1001 Ta Hsueh Rd.,
Hsinchu, Taiwan 30050, R.O.C.
**ejfarn@csie.nctu.ed.tw*
†lhchen@cis.nctu.edu.tw
‡jhliou@csie.nctu.edu.tw

In this paper, we will present a method to extract slow-motion replays in a soccer game video. According to our observation, a slow-motion replay always happens as a single shot and appears between two gradual transitions. Based on this fact, a video will first be segmented into individual shots and only those shots appearing between two gradual transitions are considered as candidate shots. There are two kinds of slow-motion replays. One comes from a standard camera and consists of some repeating or inserted frames. The other is from a high-speed camera with larger variation between two consecutive frames. Based on these features, an extractor is then provided to check if a candidate shot is a slow motion replay. Some experimental results are given to show the effectiveness of the proposed method.

Keywords: Shot detection; repeating frame; inserted frame; gradual transitions.

1. Introduction

With the development of high-speed Internet, high capacity storage and high-ratio compression standards such as MPEG-1, -2 and -4, people are quickly drowning in a growing amount of available videos. However, people do not have enough time to watch all videos in which they are interested. Fortunately, not all of the frames in a video are worth watching. To save time, developing a method to shorten the video by keeping the valuable and important frames[1–8] is necessary. For sports videos, the most valuable and important frames are slow-motion replays. In this paper, we will focus on soccer game videos.

In general, for a soccer game video, a slow-motion replay can be generated by frames (called normal frames) captured from either a standard camera or a high-speed one. If frames are from a standard camera, the number of frames captured

†Author for correspondence.

per second by the camera is the same as the number of frames played per second; if frames are from a high-speed camera, the number of frames captured per second is greater than the number of frames played per second. A slow-motion replay from a standard camera can be generated by repeating some normal frames or by inserting some frames between two consecutive frames, called target frames. Note that the inserted frames are generated through linear/nonlinear combinations of the two target frames. A slow-motion replay from a high-speed camera can be generated by dropping some normal frames.

Some approaches for detecting slow-motion replays have been presented. Kobla *et al.*[2] proposed an algorithm to check if a video contains slow-motion replays. The algorithm uses macro blocks, motion, and bit-rate information those are accessible from MPEG video. They pointed out that a slow-motion replay can be modeled as a repetitive pattern of a nonzero number of still frames followed by a nonzero number of shift frames. By detecting this pattern in the MPEG stream, they can pinpoint sequences of the video where a replay is present. However, only considering repeating frames is not enough. Pan *et al.*[4] pointed that a slow-motion replay happens between two normal shots. They presented an algorithm by using hidden Markov model (HMM) to locate slow-motion replays. The algorithm only detects slow-motion replays with repeating/dropping frames, and it assumes that a slow-motion replay always appears between two normal shots, these two points are not always true. The reason for the first one is that a slow-motion replay may be generated by adding some inserted frames; the reason for the second is that a highlight action may be represented by a group of neighboring slow-motion replays taken from different views. In this paper, we will propose a method to extract all kinds of slow-motion replays (mentioned above) from a soccer game video.

A soccer game video consists of many individual shots. A shot changes into another shot through a transition. There are two kinds of transitions: abrupt and gradual. A slow-motion replay always happens as a single shot, which appears between two gradual transitions. Based on this fact, the proposed method will first provide a transition/shot detector to locate all transitions and shots. Then each shot between two gradual transitions is considered as a candidate shot.

For a slow-motion replay from a standard camera, it is usually generated by slowing the frame rate of the playback of the recorded event. In order to slow down the frame rate, there are two different ways to generate a slow-motion replay. One is to repeat some normal frames. For this kind of slow-motion replays, we can find a phenomenon that shows very little difference between the original and its repeating frame. The other is to insert several frames between two target frames. Since an inserted frame is generated through a linear/nonlinear combination of the two target frames, we can find a phenomenon that shows the difference between the inserted frame and a target one is smaller than that between the target frame and its neighboring normal frame.

On the other hand, for a slow-motion replay from a high-speed camera, it may not contain any repeating and inserted frames, but it will drop some normal frames.

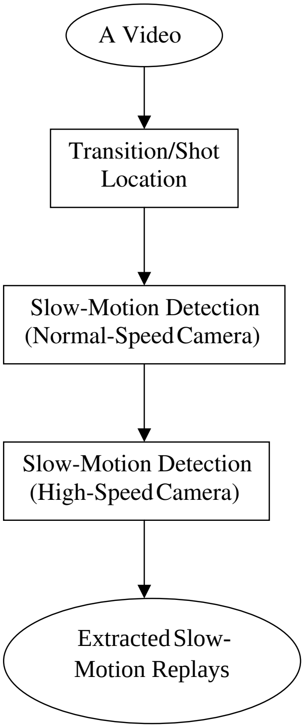

Fig. 1. The block diagram of the proposed method.

This kind of slow motion replays will have a phenomenon that shows a greater difference for most of two consecutive frames. Based on the above-mentioned three phenomena, the proposed method can then extract all kinds of slow-motion replays. The block diagram of the proposed method is presented in Fig. 1.

The remainder of the paper is organized as follows. The transition/shot locator will be presented in Sec. 2. Sections 3 and 4 will describe the proposed slow-motion replay extractor. Experimental results will be given in Sec. 5. Finally, the conclusion will be presented in Sec. 6.

2. The Transition/Shot Locator

As mentioned previously, a slow-motion replay always appears as a single individual shot between two gradual transitions. According to this fact, we will only need to check each shot between two gradual transitions to see if it is a slow-motion replay.

There are two kinds of transitions: abrupt and gradual in a soccer game video. By our observation, we find a phenomenon (called phenomenon A) that the histogram difference between any two consecutive frames in any transition is usually greater than that in any shot. Based on this phenomenon, we will provide a transition/shot locator.

In the paper, RGB color model is used, and each basic color (R, G or B) is quantized into 16 levels. For each frame n in an input soccer game video, the proposed locator first builds its color histogram $H[n][r][g][b]$. Second, for each pair of frames $n-1$ and n, the locator evaluates their histogram difference, $HD[n]$, according to the following equation:

$$HD[n] = \sum_r \sum_g \sum_b |H[n][r][g][b] - H[n-1][r][g][b]| \,. \tag{1}$$

Third, the mean (MHD) and standard deviations (SHD) of all histogram differences are computed as

$$\mathrm{MHD} = \frac{1}{N-1} \sum_n HD[n] \,, \quad \mathrm{SHD} = \sqrt{\frac{1}{N-1} \sum_n (HD[n] - \mathrm{MHD})^2} \,, \tag{2}$$

where N is the frame number of the video.

Since all transitions only occupy a small part of the whole video, MHD and SHD will approach to the mean and standard deviation of all histogram differences from shot frames. Based on this phenomenon and phenomenon A mentioned previously, for each frame n, if $HD[n] > \mathrm{MHD} + \mathrm{SHD}$, frame n will be considered as a transition frame.

Now, we will distinguish gradual and abrupt transition frames. The main difference between abrupt and gradual transitions is the frame number. An abrupt transition only needs very few frames (at most 3) to complete the shot change (see Fig. 2), this will cause the histogram difference between two consecutive frames in the transition to be larger. On the other hand, since a gradual transition usually needs more frames (see Fig. 3), the histogram difference between two consecutive frames in the transition will be smaller. Based on this fact, the locator can extract all gradual transitions. First, it computes the mean (TMHD) and standard deviations (TSHD) of all histogram differences from all transition frames as

$$\mathrm{TMHD} = \frac{1}{M} \sum_k HD[k] \,, \quad \mathrm{TSHD} = \sqrt{\frac{1}{M} \sum_k (HD[k] - \mathrm{TMHD})^2} \,, \tag{3}$$

where frame k is a transition frame and M is the number of all transition frames.

Since the total frames in abrupt transitions usually occupy a small part of all frames in all transitions, TMHD and TSHD are near the mean and standard

Fig. 2. An example for an abrupt transition.

Fig. 3. An example for a gradual transition.

deviations of all histogram differences from gradual transition frames. Based on these phenomena, for a transition frame n, if $HD[n] > \mathrm{TMHD} + \mathrm{TSHD}$, frame n is considered as an abrupt transition frame; otherwise, as a gradual transition frame. Since a shot will last at least half second, if all frames between two transition frames p and q with $q - p < 15$ are nontransition ones, the locator will reconsider these frames as gradual transition ones. Finally, all adjacent gradual transition frames will be grouped together and considered as a gradual transition, if the histogram difference between the first frame and the last frame in the group is greater than $\mathrm{TMHD} + \mathrm{TSHD}$. At this point, we can locate all of the gradual transitions. Furthermore, each shot between two gradual transitions will be considered as a candidate shot, which will be further processed to see if it is a slow-motion replay in the next section.

3. Detection of Slow-Motion Replays from Standard Camera

The frame rate of a standard camera is usually 30 frames per second, and the frame rate of a general player is also 30 frames per second. In order to turn a highlight action with normal speed into a slow-motion replay, the action frames should be extended. For example, if we want to change a 5-second action with 150 frames into an 8-second slow-motion replay with 240 frames, 90 extra frames should be added. There are two ways to reach this aim. One is to repeat some frames in the action.

The other is to insert some frames between some pairs of two consecutive frames. These two ways will be described in the following paragraphs.

The first way is to repeat some frames in the original action (see Fig. 4). Based on the fact that the difference between the original frame and its repeating frame is very small, we can extract all repeating frames. Let NT be the total number of pixels in a frame, $(R[n][i][j], G[n][i][j], B[n][i][j])$ be the color values of pixel (i, j) in frame n. Let $DF[n]$ be the difference between frames $n-1$ and n and computed as

$$DF[n] = \frac{1}{NT}\sum_i \sum_j ((R[n][i][j] - R[n-1][i][j])^2 + (G[n][i][j] - G[n-1][i][j])^2 + (B[n][i][j] - B[n-1][i][j])^2)\,. \tag{4}$$

For each candidate shot (CS), each frame n in the CS is checked. If $DF[n]$ is very small, then frame n is considered as a repeating frame. Note that a sequence of neighboring repeating frames will be called a repeating group.

The second way is to insert several frames between some pairs of two consecutive frames, called target frames (see Fig. 5). Let frames p and q be two target frames, and frames $p+1, p+2, \ldots, q-2, q-1$ be inserted frames. Since each inserted frame k is a linear/nonlinear combination of frames p and q, $DF[k] < DF[p]$ and $DF[k] < DF[q+1]$. Furthermore, from our observation, we find a phenomenon that there exists a frame n between frames p and q such that the following two inequalities are satisfied.

$$DF[p] > DF[p+1] > \cdots > DF[n]\,, \tag{5}$$

$$DF[n] < \cdots < DF[q] < DF[q+1]\,. \tag{6}$$

Note that both of two sequences $(DF[p-i], DF[p-i+1], \ldots, DF[p])$ and $(DF[q], DF[q+1], \ldots, DF[q+j])$ are not monotonic, thus we can find two local maxima p' and q' near p and q. Based on this phenomenon and Eqs. (5) and (6), we can extract all inserted frames. First, for each CS, each frame n in the CS is checked. If $DF[n]$ is a local minimum, then we will search two frames p and q nearest to frame n with $DF[p]$ and $DF[q+1]$ being local maximal. That is,

$$DF[p-1] < DF[p]\,, \quad DF[p+1] < DF[p]\,, \tag{7}$$

$$DF[q] < DF[q+1]\,, \quad DF[q+1] > DF[q+2]\,, \tag{8}$$

$$p < n < q\,. \tag{9}$$

After frames p and q are found, the following two conditions are checked

$$DF[p] > 2 * DF[n]\,, \quad DF[q+1] > 2 * DF[n]\,. \tag{10}$$

If the above two conditions are satisfied, frame $p+1$ to frame $q-1$ are considered as a group of inserted frames. Figure 5 shows an example of inserted frames.

380

381

382 (the repeating frame of 381)

383

384

385

386 (the repeating frame of 385)

387

Fig. 4. Two examples of repeating frames with the number under each image indicating frame number.

404

405

406 (p) (local maximum)

407

408 (local minimum)

409

410 (q) (local maximum)

411

Fig. 5. An example of inserted frames with target frames 406 and 410.

After all frames in a CS are checked, we can obtain the number of repeating groups (RN) and the number of insertion groups (IN). If the following inequality is satisfied

$$RN + IN > 10\,, \tag{11}$$

the CS is considered as a slow-motion replay.

4. Detection of Slow-Motion Replays from High-Speed Camera

A high-speed camera has the ability to capture more frames than a standard camera in a second, thus if a slow-motion replay is from a high-speed camera, no repeating or inserted frames are needed, but some normal frames may be dropped. For illustrative convenience, a slow-motion replay for a highlight action captured by a standard camera is called a SMRS; and a slow-motion replay for an action captured by a high-speed camera is called a SMRH. From our observation, we find that a SMRH will appear in a group of adjacent slow-motion replays or appear between two normal shots. Note that when a group of adjacent slow-motion replays exists, all slow-motion replays in the group usually describe the same action with different views (see Fig. 6), and there exists at least one SMRS in the group. When a SMRH appears between two normal shots, the SMRH usually appears in zoom-in view (see Fig. 7) and has larger difference between two consecutive frames. Here, we will provide a method to extract both types of SMRH. Before describing this method, we will first give some definitions.

(a)

(b)

Fig. 6. An example of two slow-motion replays describing the same action. Some frames in a slow-motion replay with (a) player 6 captured from the back and player 18 captured from the front side, (b) player 6 captured from the front side and player 18 captured from the back.

Fig. 7. An example of a slow-motion replay appearing between two normal shots.

Definition 1. For each candidate shot s containing frames i to j, the average difference, $ADF[s]$, of all pairs of two consecutive frames in the shot is defined as

$$ADF[s] = \frac{1}{j-i+1-r}\sum_{k=i}^{j} DF'[k]\,,$$

$$DF'[k] = \begin{cases} 0 & \text{if frame } k \text{ is a repeating or inserted frame}\,; \\ DF[k] & \text{otherwise}\,, \end{cases}$$

where r is the number of repeating or inserted frames in shot s.

Definition 2. The average difference, ADF, of two consecutive frames in all SMRSs is defined as

$$ADFS = \frac{1}{RN}\sum_{s} ADF[s]\,,$$

where s is a SMRS, RN is the number of SMRSs.

Definition 3. The dominant color, $DC[n]$, of frame n and the percentage, $DCP[n]$, occupied by $DC[n]$ are defined as

$$DC[n] = (DR[n], DG[n], DB[n])\,,$$

$$(DR[n], DG[n], DB[n]) = \arg\operatorname*{Max}_{r,g,b} H[n][r][g][b]\,,$$

$$DCP[n] = \frac{\sum_{i=-1}^{1}\sum_{j=-1}^{1}\sum_{k=-1}^{1} H[n][DR[n]+i][DG[n]+j][DB[n]+k]}{PN}\,,$$

where PN is the total number of pixels in frame n.

Definition 4. The main color, $MC[s]$, of shot s and the percentage, $MCP[s]$, occupied by $MC[s]$ are defined as

$$MC[s] = (MR[s], MG[s], M[s]) = \frac{1}{FN}\sum_{n} DC[n]\,,$$

$$MCP[s] = \frac{1}{FN}\sum_{n} DCP[n]\,,$$

where FN is the total number of frames in shot s.

(a) (b)

Fig. 8. Two examples of frames captured from different distances. A frame captured from (a) long distance and (b) closer distance.

Fig. 9. An example of a close-up shot.

Note that in a soccer game video, each slow-motion replay will have green color as its main color. When a shot is captured at a closer distance, the percentage occupied by green colors is smaller and the difference between two consecutive frames will usually be larger (see Fig. 8). This characteristic will be used later. Thus, the difference between two consecutive frames in a close-up shot (see Fig. 9) is larger; this feature is the same as that in a SMRH and so the close-up shot may be considered as a SMRH. Fortunately, since the main color of a close-up shot is not green; we can use the main color to avoid this problem.

Now, we will describe the proposed method for SMRH extraction. The method contains two phases. Phase 1 will extract each SMRH appearing between two normal shots; Phase 2 will extract SMRH appearing in a group of slow-motion replays.

In Phase 1, for each candidate shot s, which is not a SMRS, we will first check $ADF[s]$. If $ADF[s] \leqq ADFS$, the shot is not a slow-motion replay; otherwise, we will find the SMRS sn with the percentage difference between $MCP[sn]$ and $MCP[s]$ being minimum, from all SMRSs. Finally, the following condition is checked

$$|MR[s] - MR[s]| < 2\,, \quad |MG[s] - MG[sn]| < 2\,, \quad |MB[s] - MB[sn]| < 2\,. \tag{12}$$

If the above condition is satisfied, shot s will be considered as a SMRH. Note that condition (12) is used to avoid a close-up shot being considered as a SMRH.

In Phase 2, for each SMRS, each of its neighboring shots s is checked. If shot s is a candidate one, we will first find the SMRS sn from all SMRSs with the percentage difference between $MCP[sn]$ and $MCP[s]$ being minimum. Then condition (12) and the following condition are checked

$$ADF[s] > ADF[sn]\,. \tag{13}$$

If these conditions are not satisfied, the process is stopped. Otherwise, the shot s will be considered as a slow-motion replay, and the same process is applied to the neighboring shot of s to see if it is a slow-motion replay.

5. Experimental Results

In our experiments, seven 5-minute video clips captured from three different soccer games are used as testing data. The result of gradual transition extraction is shown in Table 1. From Table 1, we can see that the percentage occupied by transitions and that by abrupt transitions relative to gradual transitions are small, this meets our assumption mentioned previously. All gradual transitions are detected successfully; this guarantees that all shots of slow-motion replays will be considered as candidate shots. In video clip 5, six false gradual transitions are detected; most of them are from frames with one person occupying most of the frame (see Fig. 10). These false detections will not affect the result of slow-motion replay extraction; it only raises the time spent in candidate shot checking. Fortunately, the percentage occupied by these false gradual transitions is small.

Table 2 shows the result of the slow motion detection. From Table 2, we can see that all slow-motion replays are extracted successfully. We also find that some

Table 1. The result of gradual transition extraction.

Video Clip	No. of Total Frames	No. of Abrupt Frames and Percentage	No. of Actual Gradual Transitions and Percentage	No. of Detected Gradual Transitions and Percentage	No. of Missed Gradual Transitions	No. of False Gradual Transitions Detected and Percentage
1	9256	38(0.4%)	31(4.7%)	31(4.7%)	0	0
2	9172	44(0.5%)	39(5%)	39(5%)	0	0
3	9288	43(0.5%)	12(1.5%)	15(2.2%)	0	3(0.7%)
4	9292	36(0.4%)	5(1%)	6(1.2%)	0	1(0.2%)
5	8554	23(0.3%)	6(1.1%)	12(2.3%)	0	6(1.2%)
6	9073	49(0.5%)	37(4%)	37(4%)	0	0
7	9023	40(0.5%)	28(3.3%)	28(3.3%)	0	0
Average percentage occupied		0.44%	2.8%	3.2%		0.3%

Fig. 10. An example of false detection of gradual transition.

Table 2. The result of slow-motion replay extraction.

Video Clip	No. of Actual Slow-Motion Replays	No. of Extracted Slow-Motion Replays and Occupied Percentage	No. of Missed Slow-Motion Replays	No. of False Slow-Motion Replays Detected and Occupied Percentage
1	8	9(10.73%)	0	1(0.95%)
2	7	10(13.13%)	0	3(1.92%)
3	4	4(5.43%)	0	0
4	3	3(7.41%)	0	0
5	1	1(1.41%)	0	0
6	5	7(13.85%)	0	2(5.95%)
7	5	7(15.09%)	0	2(1.24%)
Average percentage occupied		9.58%		1.44%

false slow-motion replays are extracted. This is acceptable since they only occupy a small percentage.

Our experimental results show that the proposed method actually works very well for extracting slow-motion replays from a soccer game video.

6. Conclusions

In this paper, we have presented an effective method for extracting slow-motion replays in soccer videos. As mentioned previously, a slow-motion replay always appears between two gradual transitions. To save the processing time, only those shots between two gradual transitions are required to be checked. The proposed method has provided a gradual transition detector, which can detect all the gradual transitions successfully. There are two kinds of slow-motion replays: SMRS and SMRH. Based on the features of repeating and inserted frames in a SMRS, an extractor has then been provided to successfully extract all SMRSs. Finally, a SMRH extractor is presented to extract all SMRHs by using some phenomena in a SMRH mentioned previously. Experimental results have shown that all gradual transitions and slow-motion replays can be extracted successfully.

The proposed method is very useful for both professional and general users. It can also be applied in interactive browsing, searching, indexing and skimming.

Acknowledgment

This research was supported in part by the National Science Council of R.O.C. under contract NSC-91-2213-E-009-096.

References

1. N. Babaguchi, "Towards abstracting sports video by highlughts," *Proc. IEEE*, 2000, pp. 1519–1522.
2. V. Kobla, D. DeMenthon and D. Doermann, "Detection of slow-motion replay sequence for identifying sports videos," *Proc. IEEE 3rd Workshop on Multimedia Signal Processing*, 1999, pp. 135–140.
3. V. Kobla, D. DeMenthon and D. Doermann, "Identification of sports videos using replay, text, and camera motion features," *Proc. SPIE Conf. Storage and Retrieval for Media Databases*, Vol. 3972, January 2000, pp. 332–343.
4. H. Pan, P. van Beek and M. I. Sezan, "Detection of slow-motion replay segments in sports video for highlights generation," *Proc. 2001 IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Vol. 3, 2001, pp. 1649–1652.
5. D. Rees, J. I. Agbinya, N. Stone, F. Chen, S. Seneviratne, M. de Burgh and A. Burch, "CLICK-IT: interactive television highlighter for sports action replay," *Proc. IEEE Int. Conf. Pattern Recognition*, Vol. 2, 1998, pp. 1484–1487.
6. V. Tovinkere and R. J. Qian, "Detecting semantic events in soccer games: towards a complete solution," *Proc. ICME 2001*, Tokyo, Japan, 22–25 August, 2001.
7. D. Yow, B. L. Yeo, M. Yeung and G. Liu, "Analysis and presentation of soccer highlights from digital videos," *Proc. Asian Conf. Computer Vision*, 1995.
8. P. Xu, L. Xie, S.-F. Chang, A. Divakaran, A. Vetro and H. Sun, "Algorithms and systems for segmentation and structure analysis in soccer video," *IEEE Int. Conf. Multimedia and Expo*, Tokyo, Japan, 22–25 August, 2001.

E.-J. Farn is now a senior student in the Department of Computer Science and Information Engineering at the National Chiao Tung University.

His current research interests include video compression and image processing.

L.-H. Chen received the B.S. degree in mathematics and the M.S. degree in applied mathematics from National Tsing Hua University, Hsinchu, Taiwan in 1975 and 1977, respectively, and the Ph.D. degree in computer engineering from National Chiao Tung University, Hsinchu, Taiwan in 1987.

From August 1977 to April 1979, she worked as a research assistant in the Chung-Shan Institute of Science and Technology, Taoyan, Taiwan. From May 1979 to February 1981, she worked as a research associate in the Electronic Research and Service Organization, Industry Technology Research Institute, Hsinchu, Taiwan. From March 1981 to August 1983, she worked as an engineer in the Institute of Information Industry, Taipei, Taiwan. She is now a Professor in the Department of Computer and Information Science at the National Chiao Tung University.

Her current research interests include image processing, pattern recognition, video compression, and multimedia steganography.

J.-H. Liou graduated with a B.S. degree from the Department of Electrical Engineering, Taiwan University in 1970. He acquired his Ph.D. degree from the Department of EE&CS, University of California at Berkeley in 1975. He is currently a professor emeritus associated with the Department of Computer Science and Information Engineering at the National Chiao Tung University.

His main research interest is information security.

This article has been cited by:

1. Po-Chyi Su, Chi-Heng Lan, Chin-Song Wu, Zi-Xin Zeng, Wei-Yu Chen. 2013. Transition effect detection for extracting highlights in baseball videos. *EURASIP Journal on Image and Video Processing* **2013**:1, 27. [CrossRef]
2. Vahid Kiani, Hamid Reza Pourreza. 2012. An Effective Slow-Motion Detection Approach for Compressed Soccer Videos. *ISRN Machine Vision* **2012**, 1-8. [CrossRef]
3. GUO-SHIANG LIN, MIN-KUAN CHANG, SHIEN-TANG CHIU. 2009. A FEATURE-BASED SCHEME FOR DETECTING AND CLASSIFYING VIDEO-SHOT TRANSITIONS BASED ON SPATIO-TEMPORAL ANALYSIS AND FUZZY CLASSIFICATION. *International Journal of Pattern Recognition and Artificial Intelligence* **23**:06, 1179-1200. [Abstract] [References] [PDF] [PDF Plus]