Event Detecting and Indexing in Sport Video

Dong-Liang Lee [1] and Lawrence Y. Deng [2]

[1]Dept. of Management Information System, St. John's University

[2]Dept. of Computer Science and Information Engineering, St. John's University

Email: {[1]lianglee, [2]lawrence}@mail.sju.edu.tw

Abstract-*Video content management has become more and more important now. Video content management technologies include shot detection, shot classification, special event detection, event mining, and indexing. By way of observing the change of the different shots, some important shots or even video segments should be recognized automatically. For example, close shot, slow motion shot, repeat segment all that people interest. Those shots/segments should be detected and indexed. In this paper, we proposed an integrated video content management system and tried to represent the event detecting and indexing methods for sport video. The proposed approaches have been tested in several football games. Our experimental results have indicated certainly accuracy and efficiency. We believe some more other sport videos are also can be to implement with.*

1. Introduction

The multimedia video was more popular in our digital life currently. It contained enormous and complicated data. Therefore, it was an important issues how to manipulate that intricate information effectively and how to cut down time for consuming efficiently.

Due to the difficulty in the management of the video content, it would be much appreciated for the help of the video managing if we could established a system which could search the spectacular video fragments, such as the slow-motion video segment, replay video segment or the shot-switching segment.

2. Related Works

2.1 The Video File Format

The composition of a video film was the frame basically, and the photogene of human eyes was the main factor of a video media essentially. The human eyes could not aware the time-interval between two frames by playing a series of video frames quickly and continuously, such as 30 frames per second. If we used 3 Bytes for color information and the image size was 800 x 600, and the storage memory space of 2 hours video program could be taken up to 311GBytes approximately (30 x 7200 x 800 x 600) [4].

2.2 Shot Detection

The first step to process a multimedia video was shot detection generally. Thus, the accuracy of the shot detection would be effect to the follow-up processing absolutely. In this paper, we cut apart in several shots with shot detection technology which tried to close to the technique of a cameraman in a real world, for example, a sport video program, a cameraman would change the shot to emphasizing a sport star in the field. Therefore, how to separate a changing video shot was a primary semantic process practically. Following were some common algorithm of shot detection [5,6,7,8,10] :

(1) Comparison of Pixels

(2) Comparison of Block Area

(a)Likelihood ratio

(b) x^2 detection

(3) Comparison of Histogram

2.3 Shot Classification

The shot classification was a cluster effect of multimedia video films. A. Ekin proposed a Golden Section by separating a series of video frames of soccer video program, he also considered the characteristic of the soccer field, and created the different shot through the ratio of frames separating, which was mentioned as the Robust Dominant Color Region Detection expectedly. He utilized the main color for the basis of shot changing in whole video program, it classified to global vies, zoom-in and close-up. It was easy to express a high level meaning with low level of color detection in the video frames simultaneously [1,2,3].

2.4 Algorithms of the dynamic analysis of video

The common analysis algorithms of a video were listed following: Static background subtraction, Temporal differencing, Vector model, and Wronskian variation detection. [11]

For the purpose to enjoy those impressive pictures, the slow motion frames and re-play frames were utilized occasionally. And these video frames took a great amount in a sport video program frequently.

3. Method of key Frame Grabbing

We proposed two parts for the key frame grabbing. First, we integrated the algorithm of the dynamic analysis of video previously. Second, we used the different pixel of pictures for grabbing the dynamic variation of each frame.

The different pixel of pictures defined the current frame as f_i , and f_{i+1} for the next frame. We could calculate the different pixel value between f_i and f_{i+1}, and recorded the time period for manipulation later if the different value was higher than the threshold value. Thus, we could compare the structure of neighbor two frames by pixel alternatively.

3.1 Experimental Environment

In this study, we used an AVI format of one game from FIFA World Cup series. The process of the image processing and grabbing are shown in the Fig. 1.

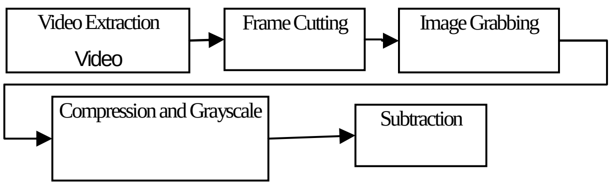

Fig. 1: Diagram of Image Processing

3.2 Video Frames Cutting

We extracted the whole video film into several frames for analyzing processing consequently which was transformed to 30 frames per second.

3.3 Video Frames Grabbing

Following steps were the steps or methods on the comparison of images: Image Compression, Gray scaling, Subtraction and Differentiation.

3.4 Gray Scaling

Gray scaling was defined as the brightness of an image. It was measured separately in some color model instead of the RGB mode, such as HSI, YUV, and YCbCr. Thus, for the operation on the brightness, we need to make a translation to the gray scaling of an image. In the HSI model, the Brightness *I* was calculated by (R + G + B) / 3. We also considered the other Brightness *Y* in the YCbCr model which is shown as below:

$$\boldsymbol{Y = 0.299 * R + 0.588 * G + 0.114 * B} \quad (1)$$

which X,Y,Z axis were indicated with R, G, B value, then the gray scaling *Y* would be the small value between the RGB(r, g, b) value of the BMP picture and the coordinates of (Y,Y,Y).

3.5 Comparison with Subtraction

The subtraction of two pixels was defined as the absolute value of the subtracting by the two gray scaling value Y of the opposition point between two continuous frames, which shown as following:

$$Difference = \left\| \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nn} \end{bmatrix} - \begin{bmatrix} B_{11} & \cdots & B_{1n} \\ \vdots & \ddots & \vdots \\ B_{n1} & \cdots & B_{nn} \end{bmatrix} \right\| \quad (2)$$

4. Semantic Analysis of Video Contents

According to the process flow-chart of the image processing, we could define the description of the shot in a video frames. There were some characteristics of grabbing portion, such as classification of shot, moment of shot, detection of replay video and slow motion video. Followings were detailed discussion for each method of detection.

4.1 Shot Classification

In the video shot of a football game, we could separate into Long Shot and Close-up Shot. The differentiation of these two shot was based on the variation value of two continuous video frames.

4.2 Process of Shot Classification

A shot was defined as a series of video frames from a single video camera. It also had a relation between the time and space in the video frames of same shot respectively.

The steps of shot research were shown as followings: (1) Video film importing. (2) Cutting into several individual frames from the importing video. (3) Setting the critical variation value of image pixel. (4) Calculating the movement value of each frame as the searching method. (5) Completing the detection of imported video film with the detection method of slow motion and replay. And determining the shot change with the comparison of pixel that discussed on section 2.2. (6) Finally, grabbing out the video contents after the comparison procedures, such as replay, slow motion, and shot change.

4.3 Semantic Theory of Video Contents

We could produce a histogram by the moment of video contents with the variation value shown on section 3.3 previously. Furthermore, we also calculated the possible moving model when the shot change, replay or slow motion occurred.

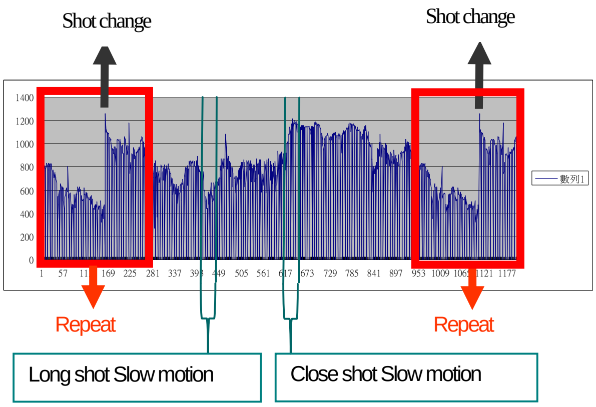

Fig. 2: Movement Analysis of a Video Content

Therefore, we could summary a shot change occurred when the variation value of movement was too large which shown on Fig. 2 obviously. There were two blocks on the right and left side of Fig. 2, those were present 'Repeat' segments. And those 'Repeat' segments also could be the same and continuous video moment instead. We also found the central part of Fig. 2 were existed slow motion frames. And we could detect it was 'Long Shot' or 'Close Shot' by a period of continuous video moments.

4.4 Method of Shot Detection

The cut shot could be detected by the Comparison of Pixels. The cut shot was an effect resulting from connecting of two independent video fragments which is shown in Fig. 3 respectively. We also found that the variation value of video movement raised or dropped tremendously when the cut shot occurred simultaneously. Therefore, we set a threshold value for the diagnosis of cut shot situation [9].

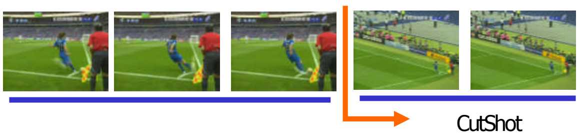

Fig. 3: A cut shot of video frames

4.4.1 The Variation Value of Pixels (Threshold Value)

The threshold value should be set previously, such as 75 for the default value. We could compare the pixel variation value of video frames. If he subtract value of each pixel was greater than the threshold value we marked it with white, and black for the less contrariwise.

4.5 Detection of Replay Frames

For the detection of replay frames, we named a series number for a same and continuous video frames repeatedly. If the numbers of a series of video frames ($A_1 \ldots A_n$) was same as another frames ($B_1 \ldots B_n$), we could indicated these two video frames were repeated apparently. And we could detect the replay video frames undisputedly.

Thus, we transformed the video frame to a matrix shown on section 3.2, and recorded the continuous values of each fragment with previous method. Finally, we checked those continuous frames were identical or not with comparison of the existed database. And we could make sure that the replay frames was detected if those frames were same coincidently.

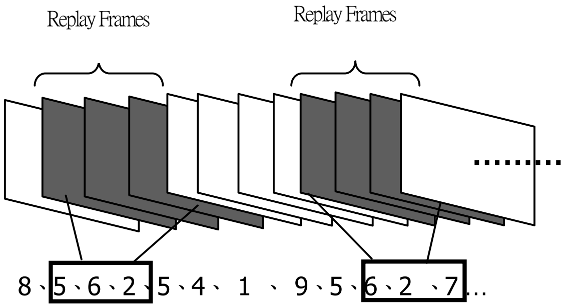

Fig. 4: A Concept Diagram of Replay Frames

4.6 The Detection of Slow Motion Shot

We had analyzed the movement of a period of slow motion video, and found the close shot of slow motion vibrated within a higher range area (shown as Fig. 5). On the contrary, the long shot of slow motion would occur around lower range area. These were much help for us to consider the regular pattern of slow motion easily.

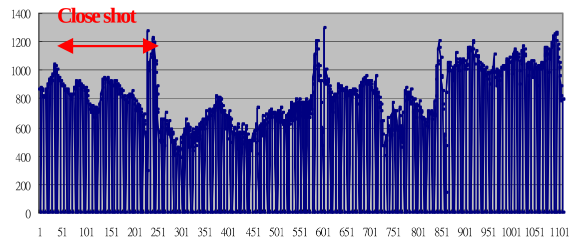

Fig. 5: The Movement Analysis of Slow Motion Shot

4.6.1 Detection of Close Shot of Slow Motion

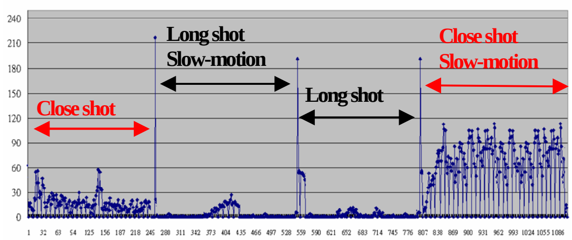

Fig. 6: The Threshold Control of Movement Analysis of Slow Motion Shot

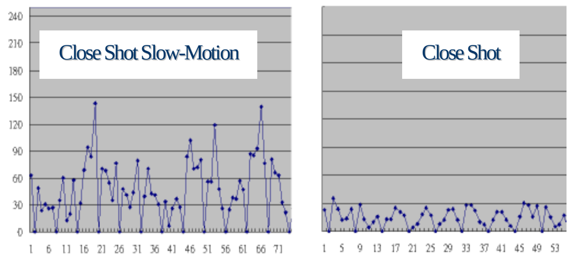

Fig. 7: The Analysis of Close Shot

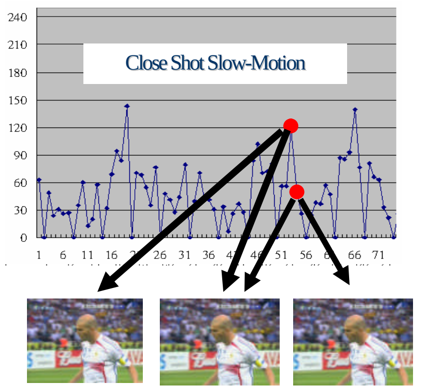

Fig. 8: Searching for the Objective Frames

According to the characters that discussed in the previous section, we could enlarge the higher range area of close shot with the control of variation threshold value. The value of the long shot and the slow motion long shot would be approached to 0 evidently (Fig. 6). We also found that the vibration margin range was much wider and obvious in close shot with same analyzing procedure (Fig. 7). There was a sharp peak in the diagram when slow motion video occurred in a series of higher similar video frames. And compared with the original video frames, we could get three similar video frames when the vibration range dropped unexpectedly (Fig. 9). We marked it with a threshold value (Fig. 8). Therefore, we could make sure that it was a close shot of slow motion video frames if those video frames matched the threshold value for a continuous and similar video fragment actually.

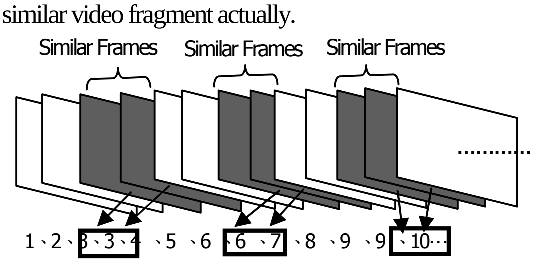

Fig. 9: Diagram for a Continuous Frame of Slow Motion Video

4.6.2 Detection of Long Shot of Slow Motion

We removed the zero movement of AVI coded video of long shot as shown in Fig. 5, the result was shown as Fig. 10 and Fig. 11. There was a little difference between these two shot. If we set the difference value in a fixed range, it would be easy to distinguish whether it is a long shot of slow motion or not. Although it would be easy to misjudge for a slow motion from a regular video frames by smaller movement and range area. We still could tell the long shot slow motion from the vibration range with a higher semantic value fortunately.

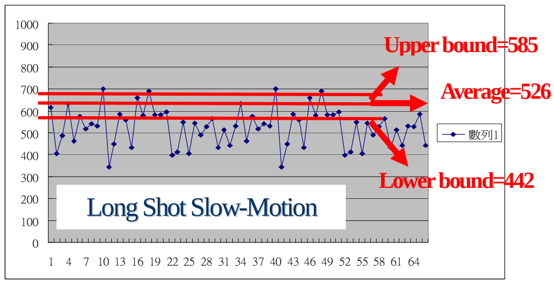

Fig. 10: Diagram of Moment Analysis for Long Shot Slow-Motion

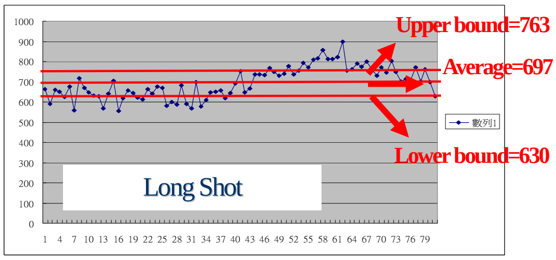

Fig. 11: Diagram of Moment Analysis for Long Shot in Normal Speed

5. Implementation

5.1 Testing and Procedures

We selected a splendid recorded video of 2006 FIFA game for the experimental video contents.

We set three parts for detection, the first were the detection of the cut-shot change, the second was the detection of replay video frames, and the last was the detection of slow-motion frames. All tested video frames were the AVI format with 320x240 size and play by the speed of 30 frames per second (fps). There were two standard evaluations to estimate our result: the "Precision" and the "Recall."

$$Recall = \frac{correct}{correct + miss} \qquad (3)$$
$$Precision = \frac{correct}{correct + false}$$

And the 'correct' mean the detected number actually, and 'miss' was the missed detected number which the shot changed should occur theoretically. And the 'false' represented the detected number by erroneous judgment which the shot changed should not existed. Therefore, we could evaluate the proportion of missed detection by the 'Recall', and the proportion of erroneous detection by the 'Precision' respectively..

5.2 The Detection and Result

The tested video was a FIFA soccer game with 36 minutes segment. We had cut shot change, Replay shot and show-motion shot on this tested film, the tested results were shown as Table 1 and Table 2.

Table 1:The Sample Data of the Testing Film

Video Film Name	FIFA Game France vs. Italy
Video Film Length	36 minutes
Frame Number	64800 frames
Shot Change	110
Replay	6
Long Shot Slow-motion	14
Close Shot Slow-motion	22

Table 2: The Comparison Data of Testing Result

Shot Type		%
Shot Change	Recall	71.6%
	Precision	93.5%
Repeat	Recall	99.2%
	Precision	87.6%
Long Shot Slow-motion	Recall	85.7%
	Precision	54.5%
Close Shot Slow-motion	Recall	92.0%
	Precision	61.0%

6. Conclusion

We could detect different events by calculating the movement subtraction of picture frames. And we analyzed the descriptor of shot ontology by the characters of each video fragment separately. We switched the shot changes by the shot classification in time relationship that conducted an inference of fragment of slow motion video or replay video concurrently. This was verified with three different video films correctly.

After the indexing with our system, those soccer video films could detect the events of shot changing precisely. We would try to enlarge our effort to detect all events in a series of sport game video content, and also enhanced the termination event's detection eventually. Especially, there was an erroneous judgment occurred easily when a man with a white shirt in the field. It made the identification difficulty by the higher difference value. Finally, we would extend our system to all kind application of sport video program generally that reduced much more of time consuming problem prospectively, such as the video content searching of TV program or movie.

References

[1] A. Ekin, et.al, "Shot Type Classification By Dominant Color for Sports Video Segmentation and Summarization", *Acoustics, Speech, and Signal Processing, Proceedings of (ICASSP '03), 2003 IEEE International Conference on Volume 3,* 6-10 April 2003, pp. III-173-6, Vol. 3.

[2] A. Ekin, et.al, "Automatic soccer video analysis and summarization," *IEEE Transactions on Image Processing,* vol. 12, no. 7, pp. 796–807, July 2003.

[3] A. Ekin, et.al, "Robust dominant color region detection and color-based applications for sports video", *Proceedings of 2003 International Conference ICIP 2003,* 14-17 Sept. 2003, pp. I-21-4, Vol. 1.

[4] Foresti, G.L.; Marcenaro, L.; Regazzoni, C.S., "Automatic detection and indexing of video-event shots for surveillance applications", *IEEE Transactions on Multimedia,* Volume 4, Issue 4, Dec. 2002, pp. 459 – 471.

[5] Ullas Gargi, et.al., "Performance Characterization Transaction on Circuits and Systems for Video of Video-Shot-Change Detection Methods" *IEEE Technology,* Vol. 10, No. 1, February 2000.

[6] R, Kasturi, R. Jain, Dynamic vision , in Computer Vision, Principles, *IEEE Computer Society Press*, Washington DC, 1991, pp. 469-480.

[7] C. M. Lee & M. C. Ip, "A robust approach for camera break detection in color video sequence", *Proceeding of IAPR Workshop Machine Vision Applications*, Kawasaki, Japan, 1994, pp.502-505.

[8] Rainer Lienhart, "Comparison of Automatic Shot Boundary Detection Algorithms" *Storage and Retrieval for Still Image and Video Databases VII 1999*, Proc. SPIE 3656-29, Jan. 1999

[9] H. C. Liu & G. L. Zick, "Automatic determination of scene changes in MPEG compressed video", *Proceeding of SPIE/IS&T Symp. Electronic Imaging Science and Technology, Digital Video Compression: Algorithms and Technologies*, Vol. 2419, 1995.

[10] N.V. Patal & I. K. Sethi, "Video shot detection and characterization for video databases", *Pattern Recognition*, 1997, Vol. 30, pp. 583-592.

[11] H.J. Zhang, A. Kankanhalli and S. Smoliar, "Automatic partitioning of video," *Multimedia Systems*, 1993, 1(1), pp.10-28.