Replay Scene Based Sports Video Abstraction

Jian-quan Ouyang[1,2], Jin-tao Li[1], and Yong-dong Zhang[1]

[1] Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, 100080
oyjq@cie.xtu.edu.cn, {oyjq, jtli, zhyd}@ict.ac.cn
[2] College of Information Engineering, Xiangtan University, Xiangtan, China, 411105

Abstract. Video abstraction can be useful in multimedia database indexing and querying and can illustrate the important content of a longer video to quick browsing. Further, in sports video, replay scene often demonstrates the highlight of the video. The detection of replay scene in the sports video is a key clue to sports video summarizing. In this paper, we present a framework of replay scene based video abstraction in MPEG sports video. Moreover, we detect identical events using color and camera information after detecting replay scene using MPEG feature. At last, we propose a three-layer replay scene based sports video abstraction. It can achieve real time performance in the MPEG compressed domain, which is validated by experimental results.

1 Introduction

Multimedia analysis and retrieval is one of the hottest issues of the information research. With the development of the artificial intelligent, communication and multimedia technology, the amount of multimedia work including digital video is vast in various fields. However, Traditional text-based information retrieval technology cannot analysis the structure of the multimedia effectively and efficiently. While multimedia analysis and retrieval can provide efficacious framework of retrieving the multimedia by extracting the lower feature and obtain the semantic content and become useful both for research and application.

Among this huge amount of visual information, the need for effective searching, browsing and indexing the videos is obvious in the computer industry and multimedia manufacturer. Fortunately, it can benefit from video abstraction. Video abstraction can be defined as a brief representation of the original video stream. The goal of video abstract is to choose the representative segment from the original long video. Yet the extraction of semantic video information is still a challenge problem. Luckily, in the live sports videos, scenes of important events or highlights repeatedly played using digital video effect or adding "logo". Usually these highlights summarize the essence and exciting actions of the video. So replay scene based sports abstraction can represent the content of the sports video.

Here we present a framework of replay scene based video abstraction in MPEG sports video. The features are directly extracted from the compressed videos. Thereby our method avoids the expensive inverse DCT computation required converting values from the compressed domain to the image domain. Furthermore, the analysis of

L. Wang and Y. Jin (Eds.): FSKD 2005, LNAI 3614, pp. 689 – 697, 2005.

this allow the macroblock (MB) type and motion vector (MV) information is simple and easy to be programmed, this allows the algorithm to be performed faster than real time video playing. The scheme can also be applicable to multiple types of sports.

The rest of the paper is organized as follows. In section 2 we review previous works related to video abstraction. In section 3, we introduce a framework of replay scene based video abstraction in MPEG sports video. In section 4, we address a new technique of detecting replay scene using MPEG feature. In section 5, we use color and camera information to detect identical events and generate replay based sports video abstract. In section 6, the experimental results with various sports video evaluate the performance of our proposed method. At last, we give conclusions of the paper and future research directions in section 7.

2 Relative Works

Some works on video abstraction have been reported. For instance, Lienhart [1] firstly presented a method of taking visual prosperity to construct a skim video that depicts a synopsis of the video sequence. But it selected the semantic contents relying on the significant visual feature such as faces, motions and verbal information.

Sports video abstraction is also an interesting topic. Li [2] propose a general framework for event detection and summary generation in broadcast sports video. Under this framework, important events in a class of sports are modeled by "plays", defined according to the semantics of the particular sport and the conventional broadcasting patterns. The detected plays are concatenated to generate a compact, time compressed summary of the original video. Obviously, it is about specific tasks including American football, baseball, and sumo wrestling. Lately, Babaguch [3] propose a method of generating a personalized abstraction of broadcasted American football video. Nevertheless, it did not verify the effectiveness of other type of sports.

Moreover, detecting replay events, often representing interesting events, can be used in video summary and content analysis. Kobla [4] used the macroblock, motion and bit-rate information to detect the slow-motion replay sequences. But it can not detect the slow motion generated by the high-speed camera. Babaguchi [5] detect replays by recognize digital video effects (DVE). The model is based on the color and motion of the gradually changing boundary between two overlapped shots. But the features are not robust and need additional computational complexity. Pan [6] detected slow-motion to determine the logo template, then located all the similar frames in the video using the logo template. Finally the algorithm identified segments by grouping the detected logo frames and slow-motion segments. However, the algorithm cannot accurately detect the slow-motion replays generated by a high-speed camera, or slow-motion replays in content whose fields are sub-sampled during encoding. Y. Yasugi [7] proposed a method for detection of identical events by analyzing and matching of the live and replay scenes for broadcasted video of American football, but it ignored zoom motion. Farn [8] proposed two kinds of slow-motion replays detection method. 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. And yet its experiments mainly are validated on soccer game videos.

The main drawback of the methods above is short of generality and hard to be applicable to other types of sports. Our solution is to make use of replay scene and camera information to generate video abstraction.

3 Framework of Replay Scene Based Video Abstraction in MPEG Sports Video

Here we propose a framework of replay scene based video abstraction in MPEG sports video, which is illustrated in Fig.1. At first, we identify the replay boundary using MPEG feature including macroblock (MB) and motion vector (MV) that is easy extracted from MPEG video, then modify the result of replay boundary detection and recognize replay scenes. Moreover, we use color and camera information to detect identical events. Finally, we introduce a scheme of three-layer replay scene based sports video abstraction**.**

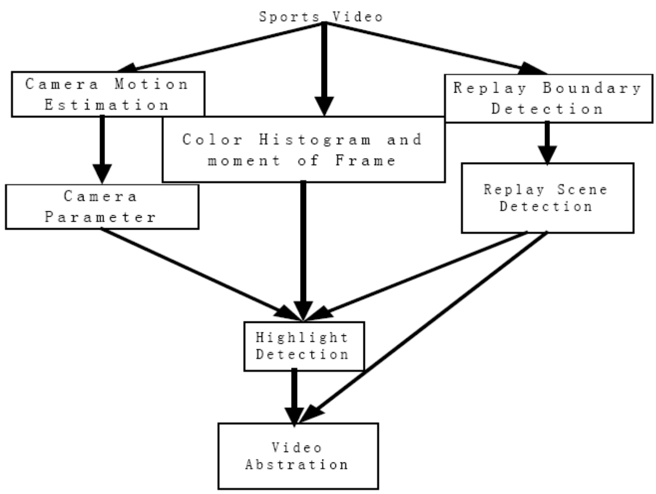

Fig. 1. The flowchart of the proposed framework

4 Replay Scene Detection

In paper [9], we present a model of replay boundary detection, and address a technique of identifying the replay boundary directly from MPEG compressed domain. It uses MPEG feature including macroblock and motion vector that is easy extracted from MPEG video, and then applies the rule of macroblock and motion vector to the detection of replay boundary.

For the reason that it is difficult to distinguish the common gradual change and logo transition, there still remain several false and missed results in replay boundary detection**.** On the other hand, the scheme of replay boundary detection cannot make use of the temporal information of diving video effectively. Thereby, the followed discussion is based two hypotheses.

I. The logos is symmetry in the broadcast diving video, namely there are logo both in the start and the end of replay scene.
II. The duration of the replay scene is shorter than the interval between the two replay scenes.

So we modify the result of replay boundary detection to detect replay scene based on the hypotheses. Experimental results validate the efficiency of this method.

5 Replay Scene Based Video Abstraction

5.1 Using Color and Camera Information to Detect Identical Events

Color feature is commonly used in video analysis and retrieval. In the MPEG compressed video pixel values are not available directly. The DC terms of I frames can be obtained directly from the MPEG sequence, and the DC terms of P and B frames can be reconstructed in [10].

Moreover, camera motion can reveal the semantic information in the video. In the replay scene, the same event which is often a highlight in the sports video is repeated several times, and often the scene of highlight which is captured by camera in a different perspective view is semantic identical to the replay scene.

We use epipolar line distance based outliers' detection method to estimate the camera motion as a motion feature for detecting identical events. Firstly we choose key frames of the video, and compute the Euclidean distance of the key frames of the replay scene and the shots before the replay scene. If the distance is below a predefined value, then the shot is a candidate shot.

Then we recover the true motion vector by estimating the camera motion, and compute the similarity of the replay scene and the candidate shots.

Similar to [7], we calculate the average true motion vectors in both the candidate and replay shots, respectively. By comparing the average true motion vectors acquired from the candidate and replay shots, if the Euclidean distance between the replay shot and the candidate shot is lower than the threshold, the candidate shot can be recognized as the live identical shot.

The identical events detection algorithm is stated as follows.

Step 1. Estimate camera motion parameter in the MPEG compressed domain.
Step 2. Recover true motion vector,

$$x_i'' = x_i' - \frac{p_1 x_i + p_2 y_i + p_3}{p_5 x_i + p_6 y_i + 1}, y_i'' = y_i' - \frac{-p_2 x_i + p_1 y_i + p_4}{p_5 x_i + p_6 y_i + 1}$$

, where (x'', y'') is the image coordinates of recovered motion vector in two neighboring frames, (x,y) and (x', y') are the image coordinates of corresponding points in two neighboring frames, $i = 0, 1, \cdots, m-1$, m is the number of feature points.
Step 3. Calculate average value of the recovered motion vectors in the frame and shot. If both the average value and direction of the shot near to these of the replay scene, then the shot is a candidate shot.

Step 4. Select key frames based on shortest path based algorithm. If it is I frame, then directly extract the DC coefficient, else estimate the DC coefficient using the method in Yeo [11].
Step 5. Compute the Euclidean distance between the replay shot and the candidate shot. If the distance is lower than the predefined value, then the candidate shot can be determined as the identical shot.

5.2 Replay Scene Based Video Abstraction

After detecting the replay scene and linking up live and replay scenes, video abstraction can build on connecting the highlights correspond to replay scene and live scene. We introduce three types of summaries:1) The key frames of the replay and live shot, 2) all replay scenes in a sports game, 3) all live scenes in the same game. The first type of summary is a still-image abstract, the last two type of summaries are moving-picture abstract. Because the users may want to a quickly preview of the video owing to the limited bandwidth, such as in a wireless network, the key frames is the preference. The users can also select the live or replay scene for the rich details of the game.

Furthermore, we propose three-layer replay scene based sports video abstraction. The top layer is the representative scene, which can illustrate the lifespan of the key

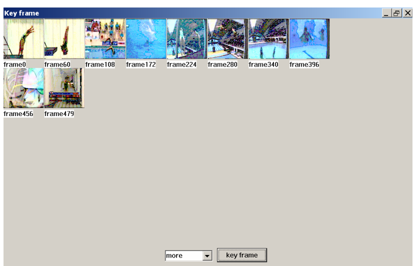

Fig. 2. Key frame interface

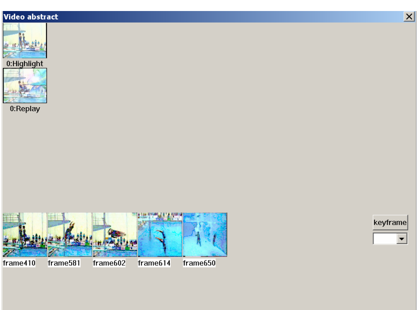

Fig. 3. Representive shot interface **Fig. 4.** Representative scene interface

actions in the sports video; the middle layer, namely the representative shot, organizes the scene summary; the bottom layer is constituted of key (or representative) frames. Moreover, as shown in Fig.3., the corresponding key frames of replay shot or live highlight is arranged to the bottom of the screen to show the detailed information. When user clicks the replay shot or live highlight, the relevant key frames would be shown. The hierarchical replay based sports video summary at the key frame layer is as shown in Fig.2, representative shot and representative scene interface is also as shown in Fig.3 and Fig.4, respectively.

To present and visualize the sports content for summarization, the representative scene, representative shot and representative frame can express the video abstraction in various granularities. Among the three-hierarchy summarization, representative scene can convey the comprehensive semantic meaning of the sports video, representative shot can demonstrate the exciting action in the video, while key frame address the detailed information. So these three hierarchical summaries can express the various video content in increasing granularity.

6 Experimental Results and Analysis

6.1 Replay Scene Detection

The test data is a set of sports video clip from the live broadcasted TV. They are "The 9th FINA Swimming Championships Fukuoka 2001" including the "3m Synchronized Diving Man (A1)" , "10m Synchronized Diving Women (A2)", "3m springboard Diving Men (A3)" "3m springboard Diving Women (A4)" ,"10m

Table 1. The experimental results of replay scene detection

	A1	A2	A3	A4	A5	A6	A7	B1	B2
Video length	46:54	44:13	1:25: 9	1:9:24	1:26:43	49:3	1:9:10	1:1:35	50:37
Total replays	40	40	72	60	72	40	55	36	32
Detect	40	38	66	52	65	40	52	26	25
False alarm	0	0	0	0	0	0	0	0	0
Miss	0	2	6	8	7	0	3	10	7
Recall	100%	95%	91.7%	86.7%	90.3%	100%	94.5%	72.2%	78.1%
Precise	100%	100%	100%	100%	100%	100%	100%	100%	100%

Table 2. Summary of results of table1

	average	The best	The worst
Recall	91.2%	100%	72.2%
Precise	100%	100%	100%

platform Diving Men(A5)" , "3m Synchronized Diving Women (A6)","10m platform Diving Women(A7)"; "The 17th FIFA world cup 2002" including the "Brazil vs. Turkey(B1,B2)". Total length of the test MPEG-1 video clip is 9:22:48, each frame is 352x288 pixels in size, and the frame rate is 25 frame/s.

Table 1 lists the experimental results of replay scene detection, and table 2 summarizes the results.

The precise of replay scene detection is desirable, but the recall depends on the result of the method of detecting replay boundary [9], the method of replay boundary detection should be improved further.

For the highlights in the sports video often can be replayed in slow motion, Kobla [4] used the macroblock, motion and bit-rate information to detect the slow-motion replay sequences. But it is only effective for detecting still frames of the replay sequences. Our method can detect the slow motion generated by the high-speed camera.

6.2 Detection Identical Events

We only list the experimental results of identical events detection of A1, A2 and A6. The accuracy and precise of identical events detection is shown in table 3.

Table 3. Eexperimental results of identical events detection

	A1	A2	A6
Video length	46:54	44:13	49:3
Total live scenes	40	40	40
Detected	36	29	36
False alarm	4	6	4
Miss	4	11	4
Recall	90%	72.5%	90%
Precise	90 %	82.8%	90%

As shown in table1, table2 and table3, the accuracy and precise of replay scene detection is fairly good, and the recall of identical events detection is also 100%. But there are still remain some error identical events detection. The false alarm and miss identical events detection mainly result from the approximate error of the camera motion estimation. Compared to the method of Tausig [7], our method can detect highlight by camera motion information, while Tausig ignore the zoom operation.

6.3 Evaluation of Replay Scene Based Video Abstraction

For directly working in the compress domain, the hierarchical summaries can be generated in real-time.

Also, as for the quality of video abstract, He [11] proposed four C's rules to measure the video abstract: conciseness, coverage, context, and context. Conciseness means the selected segment for the video summary should contain only necessary information. Coverage focuses on covering all the "key" points of the video. Context indicates that the summary should be such that prior segments establish appropriate context. Coherence contains the criterion of natural and fluid.

For conciseness and coverage, because replay scenes often drop a hint of the interesting or key the events in the sports video, our abstraction has sufficient in a compact form. Furthermore, the highlight and replay events express the context of sports video. At last, the highlight and replay events in rreplay scene based video abstraction are arranged in original temporal order.

7 Conclusions and Future Research

We have addressed a scheme of replay based video abstraction in MPEG compressed sports video. In sports video, replay scene often implies the emergence of highlight or interesting event of the video. So we apply microblock and motion vector information to detect the replay scene effectively. Moreover, we link up highlight and replay scene using color and camera information. Finally, we propose a three-layer replay-based sports video abstract. For working on features directly from the MPEG compressed domain, it can perform in real time. In the mean time, experiments verify the highlight extraction approach is more robust than current method.

The future work is to integrate our method into semantic sports video abstraction scheme, and apply to the Digital Olympic Project in China.

Acknowledgements

This research has been supported by National Science Foundation of China (60302028, 60473002), Beijing Science Foundation of China (4051004), Scientific Research Project of Beijing (Z0004024040231) and Scientific Research Fund of Hunan Provincial Education Department (03C484).

References

1. R. Lienhart, S. Pfeiffer, and W. Effelsberg.: "Video abstracting," Communications of The ACM, (1997)55–62.
2. B. Li, and MI.Sezan.: "Event Detection and Summarization in Sports Video", Proc. IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'01), (2001)132-138.
3. N. Babaguchi,Y. Kawai, T. Ogura and T. Kitahashi.: ``Personalized Abstraction of Broadcasted American Football Video by Highlight Selection," IEEE Trans. Multimedia,VOL.6,NO.4,AUGUST 2004,(2004)575-586.

4. V Kobla, D DeMenthon, and D. Doermann.: Detection of slow-motion replay sequences for identifying sports videos, Proc. IEEE Workshop on Multimedia Signal Processing,(1999) 135-140.
5. N Babaguchi, Y Kawai, Y Yasugi and T.Kitahashi.: Linking Live and Replay Scenes in Broadcasted Sports Video, ACM International Workshop on Multimedia Information Retrieval, (2000)205-208.
6. H Pan, B Li, and M I Sezan.: Automatic detection of replay segments in broadcast sports programs by detection of logos in scene transitions, IEEE International Conference on Acoustics Speech and Signal Processing, IV, (2002)3385-3388.
7. Y. Yasugi, N. Babaguchi and T.Kitahashi.: Detection of Identical Events from Broadcasted Sports Video by Comparing Camera Works, Proceedings of ACM Multimedia 2001 Workshop on Multimedia Information Retrieval (MIR2001), Ottawa, (2001)66-69.
8. EJ. Farn, LH Chen, and JH Liou.: A New Slow-motion Replay Extractor for Soccer Game Videos, International Journal of Pattern Recognition and Artificial Intelligence, Vol. 17, No. 8,(2003)1467-1482.
9. Jianquan Ouyang, Li Jintao, Zhang Yongdong.: Replay Boundary Detection in MPEG Compressed Video, IEEE The Second International Conference on Machine Learning and Cybernetics, Xi'an China, (2003)2800-2803.
10. B.-L. Yeo and B. Liu.: Rapid scene analysis on compressed videos, IEEE Trans. On Circuits and Systems for Video Technology, Vol. 5, No. 6,(1995)533-544.
11. L. He, E. Sanocki, A. Gupta, and J. Grudin.: Auto-summarization of audio-video presentations, in proceeding of ACM Multimedia, (1999)489–498.