Research on sports video analysis

Hui Li[1], Linghua Ran[2] and Yang Wang[1,*]

[1]Physical Education Department, Harbin Engineering University, Harbin 150000, china
[2]Physical Education College, Harbin Normal University, Harbin 150001, china
aoled@126.com

Abstract. Motivated by principles of natural language processing, we talk about rule-based methods for sports video analysis and propose a video parsing system by using grammar. Firstly, we segments video into elementary shots. Then through event detection, shots are annotated with semantic labels to form a token sequence. Finally, the sequence is parsed and validated with grammar to construct a table of content for the video.

Keywords: Data mining, Petri nets, Grid scheduling, Job scheduling model Video Analysis, Video Retrieval

1   Introduction

By automatic extraction of highlights in sports videos, it's probable to get initial information, such as splitting video contents to highlight part and not wonderful part. But it's not possible to analyze the structure of sport competition and related events, like dividing diving competition to several rounds to recognize events such as players' appearance, jumps. Here using for reference the principle and idea of natural language processing, we take rule-based method to discuss general sematic events in sports videos as well as their interrelationship.

Most of the existing analysis methods for sport videos are centered on semantic annotations such as shot classification, highlight retrieval, event detection etc. Despite some documents mentioned structural analysis of sport videos, such work refers to largely the breakdown of basic scenes. As indicated in [1-2], soccer videos include ongoing and suspended part; in [3-4], the concern is detecting serving scene of tennis match; [5-6] raised a universal algorithm with no need of domain model [7-8]. It utilized the clustering algorithm based on time constraint to create a three-tier structure of the video: shot, lens and scene. However, in sport videos, different competition is not of similar structure. Such uniform structure is not appropriate for sport videos [9].

In the paper, we propose a grammar-based strategy to analyze sport videos. It combines the function of semantic annotation and structural analysis to automatically generate videos' index and catalog. Unlike other video analyzing systems here, the proposed system is capable to discern hierarchical framework of sports match, easy for users to browse video contents. Moreover, with such structure expressed as

grammar, it can be used for error detection and recovery in the video analyzing process [10-11].

2   Sport video analysis based on grammars

From other videos like news, movies, sport videos have two unique features. The uniqueness represents the domain knowledge of sports video analysis. The first feature is there're domain-related events recurring in those videos. For audiences, such events are often the most important and meaningful fragments there. We can classify them into three types: playback events, state events and target events, as shown in Fig. 1. Playback events refer to slow-motion replayed snippets which are alternately played in the sportscast. They're indicative of fabulous moments in which audiences show interest. State events happen when the competition state has changed, e.g. scoring of diving at the end of each round; the start shot of one set in the tennis ace. The detection of state events is very significant to the analysis of video structure. Target events are some rather enjoyable specific movements in sport matches, such as players' jumps in diving competition, goals in soccer game. They're usually manifestations of kinematic relation of objects and among them.

(a) playback events   (b) state events   (c) target events

Fig. 1. sports video event

Video contents are of a variety. It's so hard to develop a universal event detection method to remove barriers between low-level features and high-level semantics. Hence we choose to talk about the application of domain knowledge, including competition-related knowledge and video production knowledge. From sports videos, we observe that:

(a)In most sport videos, in order to remind viewers, there are specific shot cuts as seen in the above picture, before and after playing back events;

(b)State events are captioned to show states of the competition;

(c)In target events, object and camera movements are noticeable; meanwhile, there are audiences' cheers and other related sounds like jumping sound in the diving and hitting sound in baseball match.

4　Experiment Design and Discussion

In order to validate analysis method based on grammar video, we implemented a diving video content analysis system. The system used Java language and Java Media Framework API in Win7, running on the Pentium IV 3.0GHz computer. Figure 3 is scene of the system operating.

When you open a diving video and analyze it with automatic analysis tools, you will get index and table on the left side of the screen. The table is of hierarchical structure, where each node is a shot with a token and start frame. Entries in the index are tokens and start frame images of every single event. You can search interesting videos or preview the whole video by using key frame images and tokens. By clicking key frames or tokens, you'll locate the video fragment, instead of pressing "Fast Forward" or "Fast Backward" for the same function. Besides, the system provides a great interactive environment. After automatic analysis, you can modify relevant analytical structure based on reported errors.

To validate the system performance, we used a great number of video data to test it. The video lasts totally over four hours. Table2 lists the data set for testing. All videos are MPEG-1 format, 352×288 DPI, FPS at 25 frames per second. These videos were extracted from different diving competitions in different venues, referring to (A) Men's Diving Synchronized 3m Springboard; (B) Women's Diving Synchronized 10m Platform; (C) Women's Diving 3m Springboard; (D)　Men's Diving 10m Springboard. All related events in in the data set were marked by professionals before the test as for the real reference standard.

Table 2. The test data set

Event type	Length of time	Replay events	State events	Target events	Total
A	0:45:56	40	24	40	164
B	0:44:16	40	25	40	165
C	1:34:24	60	125	50	245
D	1:25:44	72	150	62	294
Total	4:15:18	212	444	212	868

The system we developed has three modules: shot segmentation, semantic annotation and grammatical analysis. In specific implementation, we evaluate mainly algorithms used for respectively semantic annotation and grammatical analysis. The experiment includes two parts:

(1) Assessing the algorithm for semantic annotation, i.e. performance of event detection;

(2) Verifying the system performance in the grammatical analysis period, including recognition rate of high-level structural units and effectiveness of error reporting.

5   Conclusion

In this paper, by Natural Language Processing, we propose a new method for the analysis of sports video based on grammar. It can the comprehensive semantic annotation and analysis of the structure, and automatic generates sports video indexing and directory.

References

1. J. Assfalg, M. Bertini, C. Colombo, and A.D. Bimbo, "Semantic Annotation of Sports Videos," IEEE Multimedia, vol. 9, no. 2, 2012:78-82
2. Y.F Ma, and H.J. Zhang, "A Model of Motion Attention for Video Skimming", Proceeding of IEEE International Conference on Image Processing, Rochester, New York, September, 2008.
3. A. Hanjalic, "Generic Approach to Highlights Extraction from a Sport Video," Proceedings of IEEE International Conference on Image Processing, 2013.
4. L.Y. Duan, M. Xu, T.S. Chua, Q. Tian, and C.S. Xu, "A Mid-level Representation Framework for Semantic Sports Video Analysis,"Proceedings of ACM International Conference on Multimedia, November 2013.
5. N. Babaguchi, Y. Kawai, and T. Kitahashi, "Event Based Indexing of Broadcasted Sports Video by Intermodal Collaboration", IEEE Transactions on Multimedia, vol. 4, no. 1, March 2012:34-38.
6. L.Y. Duan, M. Xu, X.D. Yu, and Q. Tian, "A Unified Framework for Semantic Shot Classification in Sports Videos", Proceedings of ACM International Conference on Multimedia, pp. 419-420, 2008.
7. L. Xie, S.F. Chang, A. Divakaran and H. Sun, "Structure Analysis of Soccer Video with Hidden Markov Models," Proceedings of International Conference on Acoustic, Speech, and Signal Processing, May 2012.
8. D. Zhong and S.F. Chang, "Structure Analysis of Sports Video Using Domain Models," Proceedings of IEEE International Conference on Multimedia and Expo, August 2009
9. .Y. Rui, T.S. Huang, and S. Mehrotra, "Constructing Table-of-Content for Videos", ACM Journal of Multimedia Systems, vol. 7, no. 5, 1999.
10. R. Lienhart, "Comparison of Automatic Shot Boundary Detection Algorithm", Proceedings of SPIE Storage and Retrieval for Image and Video Databases VII, 1999.
11. F. Wang, J.T. Li, Y.D. Zhang, and S.X. Lin, "Automatically Extracting Highlights for Diving Video," Proceedings of China National Computer Conference, vol. 1, pp. 471-475, November 2013.