Advanced Science and Technology Letters
Vol.111 (NGCIT 2015), pp.167-170
http://dx.doi.org/10.14257/astl.2015.111.32

Research on analysis of sports video

Haitao Yang[1], Jia Wang[1] and Jingmeng Sun[2,*]

[1]Beijing University of Technology, Beijing 100000, china
[2]Physical Education Department, Harbin Engineering University, Harbin 150001, china
yanghaitao@bjut.edu.cn

Abstract. We turn to statistical-based methods and propose a statistical inference approach by using Dynamic Bayesian Network, which is able to learn automatically from data set. By soccer video analysis is as an example, the proposed method is verified by experiment. We extract the color, shape and other low-level features from the video, to detect and identify 5 kinds of high-level semantic events using dynamic Bayesian network model. The experimental results show that our method is effective.

Keywords: Video Analysis, Sports Video, Semantic Event Detection

1 Introduction

Grammar-based sport video analytical method is the one based on rules [1-2], easy for implementation and application. But it requires human setting rules or grammars for these reasons: firstly, in different competitions and photographic environments, it requires experts to set different rules as per experience, increasing difficulty in directly setting rules; next, in some sport videos [3-4], the relationship among events is not determined. It is uncertain and probabilistic association [5]. To overcome it, we'll discuss the method based on statistics. Unlike the method based on rules, the proposed method has some learning and adaption ability. Also it can take advantage of probabilistic relevance among events to improve effectiveness of event detection [6].

In previous work, some literatures investigated sport video analysis methods based on statistics. The most commonly used are Bayes Network（BN) and Hidden Markov Model (HMM), for instance, [7] used Bayes network to classify frame images in soccer videos to several typical scenes. In [8], the author introduced a sport event detection method based on HMM. However they have limitations for video analytics. Bayes network can perform quite well in classification. But it's a static classification model without capability to use fully contexts which change along with the time. HMM fits for processing time signals, like voice signal. But in video content analysis, its communication ability is restricted, primarily because video is a kind of multi-dimensional signal with both spatial information and temporal information. In light of all previous work, we introduce a more powerful sequence signal statistical tool, i.e. dynamic Bayes network (DBN) [11-12], to analyze sport video contents. The new method, on one thing, extends the modeling ability of Bayes network to sequence

ISSN: 2287-1233 ASTL

signals by considering transition probability at each moment; on the other thing, it allows to use a few state variables at a similar time point, rather than only one state variable used by HMM. Based on them, we think dynamic Bayes network is more suitable to analyze sport video contents, especially semantic events and interrelationship among them [13-14].

2 Event detection based on dynamic Bayes network

As for the detection of semantic events in video contents, it's necessary to build effective mapping relationship between low-level features and high-level semantics. Here we utilize dynamic Bayes network to create the mapping. We'll introduce how to set up dynamic Bayes network model as per domain knowledge, how to fetch effective low-level features and how to learn and infer high-level semantics.

2.1 Domain modeling

We take football videos for example to analyze the five events: shot, corner ball, free kick, progress, suspension. They all are defined by game rules according to human understandings. They have rich high-level semantics. Apparently, it's very difficult and inefficient to map directly from lower features like texture, shape and color to higher semantics. To avoid inefficiency, we suggest transforming it to an inference question, i.e. higher semantic events consisting of lower elements, which can be mapped to low-level features. In the indirect mapping mode, effective mapping will form between lower features and elements. Then by statistical inference of the relationship formed among lower elements, we can detect and recognize high-level elements. Compared with direct mapping, we think indirect mapping is more useful.

Based on the idea, as well as game rules and television relay regulations, we determine the five scenarios in football videos as lower elements. As shown in Fig. 1, they are respectively: (1) close-up shot and out-of-field shot; (2) medium shot; (3) midfield; (4) front court; and (5) penalty area. Close-up and out-of-field shots are pictures of people above the waist and audiences out of shooting site. Medium shots are pictures of one or some players in the field. Shooting scenes of midfield, front court and penalty area are long shots for different areas. In other sport videos, it's required to redefine semantic events and scene elements according to characteristics of the game.

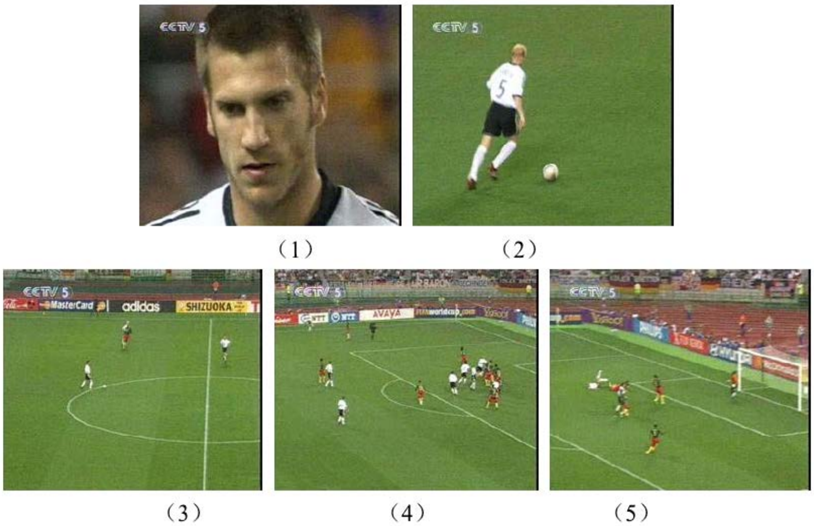

Fig.1. the basic scene in soccer video

4 Experiment Design and Discussion

The experiment has two sections. Frist of all, we evaluate the event detection effect of the method; next, considering users' requirements for automatic extraction of wonderful fragments, we regard shot, corner kick, free kick as the same fabulous fractions to assess the performance of the proposed algorithm. The testing and training data are collected from videos of four different football matches, which held in different sites and were broadcasted by different TV companies. We chose totally 54 video clips lasting from a few to over ten minutes. Before the experiment, we annotated manually all events in them to use as real reference data. Those clips constitute a video data set which lasts more than two hours. For that reason, we used half of them as training set and the rest as testing set.

Table 1. experimental results of event detection

Event	Correct	Error	Missing	Precision	Recall
Corner	25	16	2	60%	90%
Free kick	15	7	6	68%	70%
Shooting	41	16	10	71%	80%
The game proceed	116	15	11	88%	91%
Interruption of game	55	13	11	80%	84%
Total	253	68	40	78%	86%

Table 1 shows experimental results of event detection. The precision rate is on average 78% and the mean recall ratio reaches 86%. The algorithm achieved higher accuracy rate of detection in the progress and suspension of matches. In view of big content changes of them, the results are satisfactory. But it didn't do well in detecting corner ball and free kick, maybe because there's great similarity in the formation of elements of the two events. To enhance the performance, it needs finer elements and more effective algorithm for feature extraction and element recognition.

5 Conclusion

In this paper, we propose sports video analysis based on Dynamic Bayesian network statistical method. Firstly, we established a multilayer dynamic Bayesian network model based on sports video domain knowledge, including the observation layer, element layer and the event layer. In this model, the high-level semantic events is component of base, and then mapped to low-level features, so as to avoid the high-level semantics and low-level features mapping difficulties.

References

1. Wang Lian. And realize multi-mode teaching video semantic analysis. Nanjing University of Science and Technology, 2014
2. Liu Kai. Research and design feasibility basketball video tactical real-time tracking and analysis system. Capital Institute of Physical Education, 2014
3. Zhang Wei. Application Research on the technology of broadcast sports video based on Android platform. North China University of Technology, 2014
4. Ding Liwei. Study on detection method for soccer video highlights model based on HCRF. Xi'an Electronic and Science University, 2014
5. Hairihan. Research on tennis serve video event detection. Inner Mongolia Normal University, 2014
6. Wu Peng, Xinqi Lin, Li Haitao, Gao Zhi, Zhan Xiaoshan. Sports video classification based on color texture and SVM. Journal of Fujian Normal University (NATURAL SCIENCE EDITION), 2014,02:34-41.
7. The probability distribution PI said sequence starting events, A said the distribution of transfer probability between events.
8. M. Luo, Y.F. Ma, and H.J. Zhang, "Pyramidwise Structuring for Soccer Highlight Extraction," Proceedings of IEEE Pacific-Rim Conference on Multimedia, 2013.
9. 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, 2002.
10. Cui Yunxiang. Two dimensional body joint characteristics of sports video annotation based on Computer Engineering, 2014,04:252-257.
11. G. Xu, Y.F. Ma, H.J. Zhang, and S.Q. Yang, "A HMM Based Semantic Analysis Framework for Sports Game Event Detection," Proceedings of IEEE International Conference on Image Processing, 2013.
12. K.P. Murphy, "Dynamic Bayesian Network: Representation, Inference and Learning," PhD Dissertation, University of California, Berkeley, 2012.

