Advanced Science and Technology Letters
Vol.111 (NGCIT 2015), pp.161-166
http://dx.doi.org/10.14257/astl.2015.111.31

Research on multi-pattern fusion

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

[1]Beijing University of Technology, Beijing 100000, china
[2]Physical Education Department, Harbin Engineering University, Harbin 150000, china
tennis_jia@aliyun.com

Abstract. In order to effectively integrate multimodal information and multilayer constraints, we present a unified probabilistic framework for sports video analysis. Based the framework, three instances of statistical models are constructed and compared. Experimental results indicate our method with multimodal fusion processes semantic events in sports video more effectively.

Keywords: Video Analysis, Sports Video, Semantic Event Detection

1 Introduction

With the method based on statistics, we'll discuss further sport video content analysis method fusing multimode information. Semantic events in sport videos are in essence multimodal. In television relay, multimode information is integrative used to present video contents like subtitles, narrator's voices, on-site sounds, camera movements, scenarios and images etc. It is incomplete to analyze only one mode. For more effective analysis of events, it's required to study the analytical method which fuses multiple patterns. On the other hand, semantic events in those videos are not isolated. There's some logical or consequential relationship among them. In previous paper, we discussed event detection and recognition with the use of the contextual relationship based on dynamic Bayes network. Now on that basis, we'll explore how to fuse multimode information, which is a key issue we're facing here [1-2].

In recent years, the fusion of multimode information has become a hot topic in the field of sport video analysis [3-5]. Firstly we introduce the related work. Most of the multimode fusion analysis methods mentioned in previous literatures considered the detection of an isolated event. Unlike them, we propose to detect many events and analyze comprehensively the association among them.

Multi-level analysis methods based on statistics are built on the probabilistic graphical model, such as Hidden Markov Model (HMM), dynamic Bayes network (DBN) and their variants. By combining visual graphical model representation and effective reasoning and learning methods, such solutions made fairly good effects. Xie [6-7] et al. applied hierarchical HMM for unsupervised clustering to discover layered structure of video contents. Differently, we introduced the method based on dynamic Bayes network model to do the same work. Through learning of training samples, we fulfilled the detection and recognition of wonderful events in the football match [8].

ISSN: 2287-1233 ASTL

2 Multimode multi-level semantic analysis framework of sport videos

For the semantic analytics of sport videos, multimode information and multi-level constraints are important foundations to do that. But in existing papers, there isn't a solution based on statistics, which can combine them together. Here we proposed a multimode multi-level analysis framework on the basis of DBN. Then based on it, we designed three models such as: FHHMM, CHHMM and PHHMM. Firstly we'll give expression form of them based on DBN; then, we'll discuss the learning and reasoning algorithm.

Dynamic Bayes network is one kind of directed probabilistic graphical model. Its parameter can be put as (Λ, Θ). The first group Λ refers to DBN's structure-related parameters inclusive of node quantity in each frame, network topological structure etc.; the second group Θ stands for the conditional probabilistic distribution to which all connection lines in network relate, node's initial probability etc.

Fig. 1 is the graphical structure of a three-level HHMM represented as DBN. In it, Q_t^d means state variable at time point t in the level d; the total number of states in d is n_d; F_t^d is indicator variable; F_t^d=1 if and only if Q_t^d's sublevel HMM i.e. Q_t^{d+1} moves with time to the end state. What should be noted is if $F_t^d = 1$, then in every level below, there exists $F_t^{d'} = 1(d' > d)$.

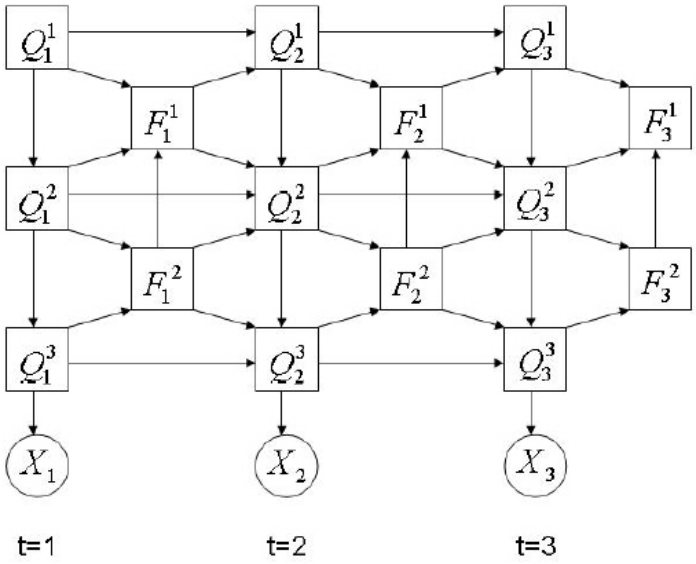


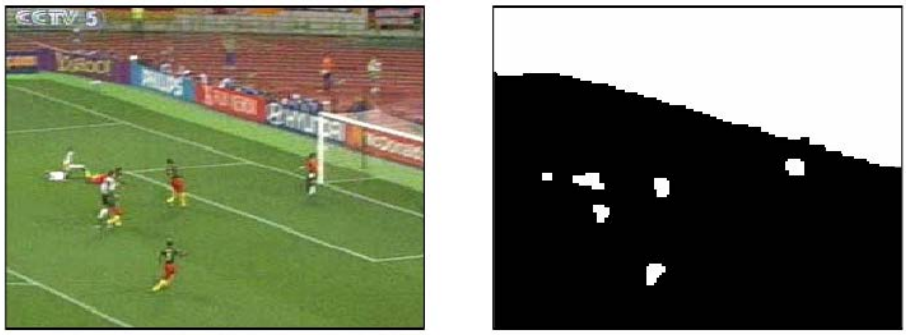
Fig.1. HMM DBN of a three layer

3 Detection of in-play and out-of-play events in football videos

To verify and compare the performance of DBN model, we applied the above models to detect “in-play” and “out-of-play” events in football videos. “In-play” events are defined as the time for the football game goes on normally. “Out-of-play” events, also called dead ball, refer to what happens when the ball passes on the ground or in the air wholly over goal line or sidelines; or when the game is called by a judge. The detection of those events is very significant to automatically generate video abstracts and make higher semantic analysis. To be specific, by detecting game progress and suspension, we can remove suspended video fragments to generate more simplified video abstracts.

Firstly we extract from video streams the color and movement features to use as observation data in different patterns. As stated in [11], those features originate from different modes, among which the correlation is very little. But at the same time, for the detection of “in-play” and “out-of-play” events, the information of mutual supplement in the two modes are valuably referential. Therefore, it requires a method which can fuse effectively modal information and utilize fully contextual restraints to detect events. Here we didn’t use audio features. That’s because they’re not quite distinguishable for the detection of “in-play” and “out-of-play” events after our observation.

[12] suggested detecting the game is ongoing or paused by according to shot classification and some heuristic rules. The method required manually setting rules and determining threshold as per experience. In [6], Xie et al. used exercise intensity and main color ratio as characteristics, with two groups of HMM to stand respectively for “in-play” and “out-of-play” events. Then based on HMM’s output probability, they segmented and recognized events by dynamic programming. On that basis, Xie developed a new method based on HHMM, which was confirmed effective by the experiment on detecting “in-play” and “out-of-play” events. Here we also implemented a method based on HHMM for event detection. We used it as standard for comparison.



(a) the original image (b) the binary image

Fig. 2. feature extraction based on dominant color

4 Experiment Design and Discussion

The experiment has two objectives:

(a) to evaluate whether our proposed multi-level DBN models like FHHMM, CHHMM and PHHMM can effectively realize fusion processing of multimode information;

(b) to compare the performance of the proposed method and others.

To fulfill the first goal, we carried out a system based on traditional HHMM as benchmark. The system uses one-mode information for processing, in other words, it applies only the color features for training and recognition. Then, it utilizes only motion features for the same task. The system contains higher and lower nodes. Nodes in higher levels correspond to "in-play" and "out-of-play" events; nodes in lower levels correspond to hidden elements. For the second goal, we implemented a system based on feature fusion as benchmark. Despite using still HHMM, this system combines motion features and color features together as observation input. It uses feature fusion method instead of decision fusion method. Unlike HHMM-based system, the proposed FHHMM, CHHMM and PHHMM create separately elements and observation for different modes. The number of different element's states affects the function of those models. In our experiment, we find whether the state number is too less (<5) or too more (>9), our models' performance is reduced. In this case, we choose the best result of each model for comparison.

The data set for testing includes 20 video clips which last from a few minutes to over ten minutes. They were chosen from five sessions of football match videos, MPEG-1 format, size 352x288, 25 frames per second. At every 0.5s, we fetch from video stream the color and motion features to decrease calculated amount. The "in-play" and "out-of-play" events in testing data set were manually remarked as real data beforehand. Cross-Validation experiment was conducted to train and assess those models. That is to say, every time 90% data is used for training and the rest for testing. Repeat ten times till all data are tested in turn.

Regarding video analysis results acquired by different models, we evaluate them based on video frames and segments. Frame-based evaluation is made to compare true data with results got by automatic analytics by frame, calculating the percentage of right marked frames versus the total. The evaluation can reflect the sensitivity of each model to local changes (Table1). There HHMM_C refers to the mere use of color features; HHMM_M refers to the mere use of motion features. Other models use the two features simultaneously.

Table 1. the experimental results based on frame

Model	HHMM_C	HHMM_M	HHMM	FHHMM	CHHMM	PHHMM
Accuracy rate	78.6	63.7	77.8	81.4	86.3	84.6

From Table1, we see PHHMM achieved the highest rate of accuracy. Compared with models using only single mode, the three multimode multi-layer DBN models obtained better results. They also performed better than HHMM using feature fusion mode. In our experiment, the feature fusion method was even worse than models

utilizing only color features. The reason may lie in too much interference in motion features. The combined application of color and motion features actually degraded the effectiveness of those features. Conversely, the proposed fusion model proved better adaptability.

5 Conclusion

In this paper, we proposed a dynamic Bayesian network based on a fusion of multi model information and multi-level constraint sports video analysis framework. On the one hand, multi-level analysis based on dynamic Bayesian network can express the domain knowledge for the topological structure of intuitive. On the other hand, the learning and inference algorithms can effectively establish a statistical interaction between multi-pattern information.

References

1. M. Han, W. Hua, W. Xu, and Y.H. Gong, "An Integrated Baseball Digest System Using Maximum Entropy Method," Proceedings of ACM International Conference on Multimedia, 2012.
2. Y. Wang, Z. Liu, and J.C. Huang, "Multimedia Content Analysis Using Both Audio and Video Clues," IEEE Signal Processing Magazine, 2010.
3. M. Barnard, J.M. Odobez, and S. Bengio, "Multi-Modal Audio-Visual Event Recognition for Football Analysis," Proceedings of IEEE Workshop on Neural Networks for Signal Processing, 2013.
4. M. Xu, L.Y. Duan, C. Xu, and Q. Tian, "A Fusion Scheme of Visual and Auditory Modalities for Event Detection in Sports Video," Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, April 2013.
5. M. Petkovic, V. Mihajlovic, W. Jonker, and S. Djordjevic-Kajan, "Multi-Modal Extraction of Highlights from TV Formula 1 Programs," Proceedings of IEEE International Conference on Multimedia and Expo, 2012.
6. L. Xie, S.F. Chang, A. Divakaran, and H. Sun, "Unsupervised Discovery of Multilevel Statistical Video Structures Using Hierarchical Hidden Markov Models," Proceedings of IEEE International Conference on Multimedia and Expo, 2008.
7. L. Xie, S.F. Chang, A. Divakaran, and H. Sun, "Feature Selection for Unsupervised Discovery of Statistical Temporal Structures in Video," IEEE International Conference on Image Processing, Barcelona, Spain, September 2013.
8. F. Wang, Y.F. Ma, H.J. Zhang, and J.T. Li, "Dynamic Bayesian Network Based Event Detection for Soccer Highlight Extraction", Proceedings of IEEE International Conference on Image Processing, Singapore, October 2004.
9. L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proceeding of the IEEE, vol. 77, no. 2, 1989.
10. T.A. Stephenson, "An Introduction to Bayesian Network Theory and Usage," IDIAP Research Report, February 2000.
11. 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 2003.

12. 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 2002.

