Advanced Science and Technology Letters
Vol.111 (NGCIT 2015), pp.156-160
http://dx.doi.org/10.14257/astl.2015.111.30

Multi-mode semantic cues in soccer video

Yu Wang[1], Yu Cao[2], Miao Wang[3] and Gang Liu [4,*]

[1]Capital Normal University, Beijing 100048, china
[2]Renmin University of China, Beijing 100048, china
[3]Harbin Sports Institute, Harbin 150000, china
[4]Physical Education Department, Harbin Engineering University, Harbin 150000, china
13488850358@126.com

Abstract. A new framework based on multi- modal semantic clues and HCRF (Hidden Conditional Random Field) for soccer wonderful event detection. Through analysis of the structural semantics of the wonderful event videos, define nine kinds of multi- modal semantic clues to accurately describe the included semantic information of the wonderful events. After splitting the video clips into several physical shots, extract the multi- modal semantic clues from the key frame of each shot to get the feature vector of the current shots, and compose the observed sequence of the feature vectors of all shots in the test video clips. Using the above observed sequence as HCRF model input in the case of small-scale training samples, establish a wonderful event detection HCRF model effectively.

Keywords: pixel unmixing, end-member extraction, data field

1 Introduction

To solve the automation problem of event detection, plenty of machine learning algorithms were widely applied, including Dynamic Bayesian Network (DBN) model, Hidden Markov Model (HMM), Conditional Random Fields model and Support Vector Machine (SVM) model [1-2]. However, those machine learning algorithm models have obvious shortcomings. To be specific, HMM requires possibly complete sample space. The model construction is complex [3-4]. The computational amount is huge. Moreover, it needs conditional independence assumption when creating the model, not describing the true structure of events [5-6]. The machine learning methods based on SVM take semantic event detection directly as feature classification problem to solve, not thoroughly utilizing semantic information, leading to bad performance of the detection. The machine learning strategies based on Conditional Random Fields (CRF) model can't describe interior structure and potential information of semantic events by defining hidden state variables when creating the event model, limiting the model's abilities to depict and recognize semantic events [7-8].

With the deeper investigation, Quattoni [9] introduced a brand new machine learning method [10-11], which is based on hidden conditional random fields (i.e. HCRF) model. The strategy incorporates merits of all algorithms mentioned above. It

ISSN: 2287-1233 ASTL

has been successfully applied in the field of gesture, voice and action recognition. The HCRF model makes full advantage of potential hidden state structure to discover effectively the internal laws of semantic events. The manifested long-distance dependence and overlapping feature are more accordant with structural features of video signals. Meanwhile, according to video multi-granularity, through overall analysis and description of video semantic events in terms of video images and audios, as well as the fusion of multimodal semantic clues and HCRF model, a feasible new framework is developed to detect wonderful episodes in football videos [12].

2 Extraction of multimode semantic clues

After analyzing the semantic structure of highlights in football match videos, it defines nine multi-pattern semantic clues, as to mine accurately the semantic information contained by such highlights and thus to express them roundly and clearly. Of the nine multimode semantic clues, the selection of scoring board SB, referee rate RR and frame motion FM are defined and retrieved by the following method:

2.1 Scoring board (SB)

When there's foul, referee will show red card/yellow card. Then on the scoring board will show the name of the offender and information relating to the red/yellow card. In this case, SB can be regarded as an important semantic clue for red/yellow episode. Here we use Harris corner detection method to separate SB Board Area (BA), defining and giving Aspect Ratio of BA (AR_BA) and Area Ratio of BA (AreaR_BA). The design formula (1) (2) and quantitative rule of semantic shots is shown as (3):

$$AR\_BA = \frac{W\_BA}{H\_BA} \tag{1}$$

$$AreaR\_BA = \frac{AreaR\_BA}{p \times q} \tag{2}$$

$$SB = \begin{cases} 1, AR\_BA(i) < T_s, AreaR\_BA(i) < T_e \\ 0, else \end{cases} \tag{3}$$

$SB = 1$ means the scoreboard shots, otherwise is non scoreboard shots. Figure 1 (a) is a representative frame of the scoreboard shot.

2.2 Referee Ratio (RR)

Referee-relating information in the football videos can be used for detecting highlights, like red/yellow incidents. Here we employ the approach in [13] to detect referee' clothing, define the aspect ratio, area rate and Aspect Ratio of MBR (AR_MBR) of the minimum bounding rectangle (MBR) in referee's clothes Area Ratio of MBR (AreaR_MBR), which are calculated by and the quantitative rule is(4) (5) (6):

$$AR\_MBR = \frac{W\_MBR}{H\_MBR} \tag{4}$$

$$AreaR\_MBR = \frac{AreaR\_MBR}{p \times q} \tag{5}$$

$$PR = \begin{cases} 1, AR\_MBR(i) < T_u, AreaR\_MBR(i) < T_r \\ 0, else \end{cases} \tag{6}$$

$PR = 1$ means the Referee shots, otherwise is non Referee shots. Figure 1 (b) is a representative frame of the Referee shot.

(a) the scoreboard lens representative frame

(b) the referee shot representation frame

Fig. 1. The scoreboard shot, representative frame of referee shot

4 Experiment Design and Discussion

Experimental videos were collected from many sessions of 2010 South Africa FIFA, 2011 EPL and 2011 UEFA, MPEG format, 352 x 288 DPI, software environment Matlab R2008a. The empirical data include training data and testing data, of which training data for corner kicks have 20 corner kick fragments and 10 non-corner kick

fragments; testing data have 30 corner kick clips and 20 non-corner kick clips; training data for penalty kicks have respectively 20 penalty and non-penalty kick fragments; testing data have 61 penalty kick fragments and 20 non-penalty kick fragments; for the red/yellow cards, training data have 20 red/yellow card clips and 10 non red/yellow card clips; testing data include 37 red/yellow card clips and 15 non red/yellow card clips. In the experiment, we use recall rate and precision rate to evaluate quantitatively the retrieval results of multimode semantic clues and detection results of above highlights. Due to the space here, Table1 only lists partial experimental video information about corner kick episodes.

Table1. Experimental video information of the corner event

Video name	ID	Matches	Date of the match	Score	Video length
South Africa World Cup	F1	England VS USA	2010.6	1:1	106
	F2	Germany VS Australia	2010.6	4:0	102
	F3	Spain VS Switzerland	2010.6	0:1	107
	F4	Germany VS Argentina	2010.7	4:1	109
UEFA Champions League	U1	Real Madrid VS dynamo Zagreb	2011.11	6:2	95
	U2	Bayern Munich VS Villarreals	2011.11	3:1	106
	U3	Napoli VS Manchester City	2011.11	2:1	100
	U4	AC Milan VS Barcelona	2011.11	2:3	101
England Football Super League	E1	Chelsea VS Wigan	2012.4	2:1	109
	E2	Manchester United VS Queens Park Rangers	2012.4	2:0	96
	E3	A Senna VS Manchester City	2012.4	1:0	107
	E4	Tottenham VS Norwich	2012.4	1:2	102

5 Conclusion

This paper presents a new framework for soccer video highlights multimodal cues and detection based on HCRF model. Firstly, the fusion of audio and video features, constructed middle level semantic space using multi modal semantic clues, make up the semantic gap from the low-level features to high-level semantics. Secondly, the

multi-pattern semantic cues to form the feature vector as the observation sequence of HCRF model

References

1. Quattoni A, Wang S, Morency L P, et al. Hidden conditional random fields. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(10): 1848-1852.
2. Yedidia J S, Freeman W T, Weiss Y. Generalized belief propagation. Advances in Neural Information Processing Systems, 2001: 689-695.
3. Ming, Ding Liwei, Maggie Jackie Chan. Soccer video highlights the fusion of HCRF and AAM detection. Journal of computer research and development, 2014,01:225-236.
4. Duan Xiping, Liu Jiafeng, Wang Jianhua, Tang dragon. A semantic level collaborative text image recognition method. Journal of Harbin Institute of Technology, 2014,03:49-53.
5. Yang Minghao, Tao Jianhua, Li Hao, nest forest at multi-channel man-machine dialogue system for natural interaction. Computer science, 2014,10:12-18+35.
6. Wang Lian. And realize multi mode teaching video semantic analysis. Nanjing University of Science and Technology, 2014
7. Tian. Study on construction of spatial knowledge obviously multi modal based on information fusion. Huazhong Normal University, 2014
8. Hu Yucheng, Yu Junqing, Huang Xianqiang, He Yunfeng, Tao. User preference mining pipe in the engine of the soccer video search. China Journal of image and graphics, 2014,04:622-629.
9. Yu Junqing, Zhang Qiang, Wang Zengkai, He Yunfeng. Using the playback scene and emotion encouragement detection in soccer video highlights. Chinese Journal of computers, 2014,06:1268-1280.
10. Zhang Yanjiao. Regional map of target detection of video abstract Gauss. Hebei Normal University, 2014
11. Lu Yafei. Study on detection of pedestrian tracking and abnormal motion video surveillance. China Jiliang University, 2014
12. Su Chenhan. Methods and annotation of video structure extraction. Computer knowledge and technology, 2014,26:6178-6180

