Exciting Event Detection in Broadcast Soccer Video with Mid-level Description and Incremental Learning

Qixiang Ye
Institute of Computing Technology of Chinese Academy of Sciences,
Beijing, 100080, China
qxye@jdl.ac.cn

Qingming Huang
Graduate School of Chinese Academy of Sciences, Beijing, 100080, China
qmhuang@jdl.ac.cn

Wen Gao，Shuqiang Jiang
Graduate School of Chinese Academy of Sciences, Beijing, 100080, China
{Wgao,sqjiang}@jdl.ac.cn

ABSTRACT

In this paper, we propose a method for exciting event detection in broadcast soccer video with mid-level description and SVM-based incremental learning. In the method, video frames are firstly classified and grouped into views in terms of low-level playfield features. Mid-level description including view label, motion descriptor and shot descriptor are then extracted to present the characteristics of a view. By using the fixed temporal structure of views, SVM classification models are constructed to detected exciting events in a soccer match. In the view classification and event detection procedures, SVM-based incremental learning method is explored to improve the extensibility of view classification and event detection. Experiments on real soccer video programs demonstrate encouraging results.

Categories and Subject Descriptors

[**Multimedia analysis, processing, and retrieval**]: – Subject 1. Multimedia Content Analysis. Subject 2. Video Summaries and Storyboards.

General Terms: Algorithms, experimentation.

Keywords: Sports video analysis, exciting event detection.

1. INTRODUCTION

Sports video programs always hold mass audience. In recent years, the amount of digitized video content has been increasing rapidly and users need to access their content through various network solutions by digital equipments. Therefore, automatic sports video analysis, for example, extracting highlights in soccer video to make it possible to deliver video clips into mobile devices or Webs, has been a coming requirement. Moreover, summarizing a long sport video into short segments will bring convenience to content-based video retrieval, and then to facilitate the users finding the preferred video content in a database. These inspire the researches on sport video summarization and highlights detection in recent years [1-11].

There are some works on general video highlight detection by replay shot detection [1], video activity analysis, audio analysis [2], structure analysis [3] and mid-level semantic extraction [4]. These works can only coarsely find highlights in sports video while most of the users look at the semantic events like "shoot on goal" in soccer, "goal" in basketball etc. For the reason that detecting events in general sports video is still an open problem at present, most of the researcher focus on a special kind of sports video, including American football [5], basketball [6], etc. in which soccer video is mostly concerned [7-12] for its high audience rating.

As the forerunner of soccer video analysis, Gong et al. [7] use player, ball, line marks and motion features to parse TV soccer programs. Xie et al. [8] proposed a method to segment soccer video into "play" or "break" segments in a HMM (hidden Markov model) framework. The low level features she adopted are video dominant color and motion activity. In [9] cinematic feature as shot type, replays and object features are integrated into a Bayesian Network classifier to identify "goal" event in broadcast soccer and basketball video. The work depends on mainly the result of replay detection result. But replay detection is still an open problem. Leonardi et al. [10] detect "goal" event in soccer video by camera motion analysis and shot boundary trigger in a Control Markov Chain framework. Detected "goal" events are then verified by the audio feature. Assfalg [11] et al. use playfield zone classification, camera motion analysis and player's position to infer highlights of non-broadcast soccer video by FSM (finite state machine). Although, good results are report, the FSM-based method need the researchers make good rules for different kinds of events by hand. Furthermore, the soccer programs from different cameraman and director hold different styles. The appearance of same soccer events may very a lot with the change of camera station. This will depress the extensibility of event detection models. For example, the event detection model built on "Europe cup" soccer programs performs bad on "England soccer" programs.

In this paper, a new method for exciting event detection methods in broadcast soccer video based on mid-level visual description and incremental SVM learning is investigated. Framework of the proposed method is described in Fig.1 (next page). For each of the video frame, we adaptively detect the playfied area based on which we extract projection profile histogram, dominant color ratio and shape features. Then SVM classifiers are employed to hierarchically classify the fames into defined views. For each view, we extract camera motion and shot boundary descriptions as the mid-level description. Finally, view label, camera motion and shot boundary descriptions together with temporal relationship are fed into a SVM classifier to identify events. In the playfied detection, view classification and event detection stages, incremental learning is adopted, which enable the method can be easily extended to soccer video of different styles. Three typical exciting events, say "shoot on goal", "goal", "placed kick", "break by offence", are selected for experiments.

MM'05, November 6–11, 2005, Singapore.

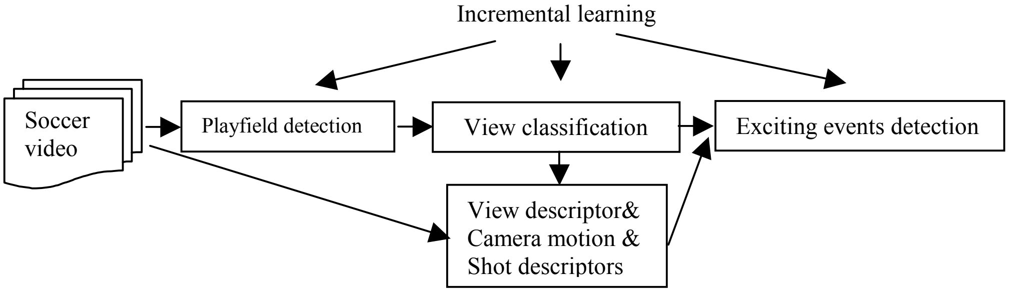

Incremental learning
Soccer video
Playfield detection
View classification
Exciting events detection
View descriptor& Camera motion & Shot descriptors

Figure 1. Framework for exciting event detection.

The rest of the paper is organized as follows. View classification and description are presented in section 2. Exciting event detection is presented in section 3. Experimental results are presented in section 4. We conclude the paper with a discussion of future work in section 5.

2. MID-LEVEL DESCRIPTORS EXTRACTION

Extracting semantics from low-level features is a traditional open problem. Building mid-level description is an effective way attempt to this problem [13]. In this work, three kinds of the mid-level description are extracted to represent the soccer video content.

2.1 Playfield segmentation

Playfield is segmented by our method in [14]. In the playfield detection process, training models are automatic build by Gaussian Mixture Models (GMMs). The model is updated online by an incremental procedure, say, automatically extracting playfield pixels as new training samples for parameters adjustments. Details can be found in [14].

2.2View labeling

For each of the soccer frame, we can semantically assign a view label to them by a hierachical classification procedure as Fig.2. There are three levels (five kinds of views) in Fig.2. The views are:“goal mouth view”, “corner view”, “middle field view”, “player close-up view” and “out-field view”. For each of the level we extract low-level features for classification.

Level-1: Playfield ratio (pr) feature to discriminate the views into in field and out field views. pr is calculated as

$$pr = playfield\ area\ /\ frame\ area \qquad (1)$$

Level-2: Discriminating “local view” and “global views” in the in-field views, and discriminate. We extract the following features to perform the classification task.

Player field ratio (pfr) (2)

Projection profile feature of non playfield (ppf) (3)

where ppf (show as Fig.3b and Fig.3d) is a 16 dimension projection on x and y direction of non-playfield area, which can reflect the playfield layout. ppf is the key feature to discriminate player close-up view and lower camera view.

Level-3: Discriminating the shapes of field in local views. Five dimensions shape features are extracted in the level. They are

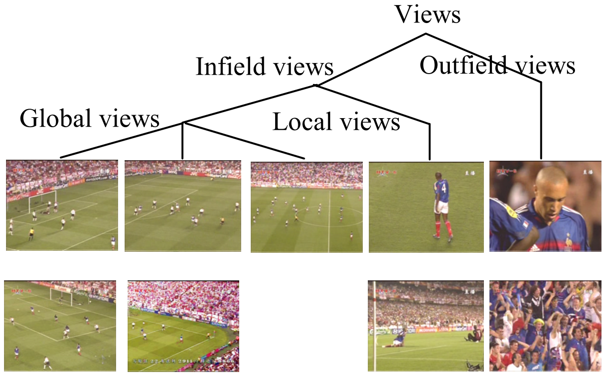


Figure 2. Hierachical view classification.

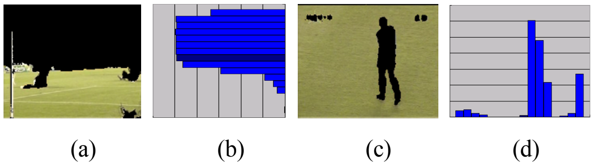


Figure 3. Projection profile feature of non playfield pixels.

ts - slope of top boundary (shown as Fig.4 a) ;

ls - slope of top left boundary (Shown as Fig.4 b);

rs - slope of top right boundary (Shown as Fig.4 b);

cp - corner position (Shown as Fig.4 b).;

bs -slope of bottom boundary(shown as Fig.4 c).

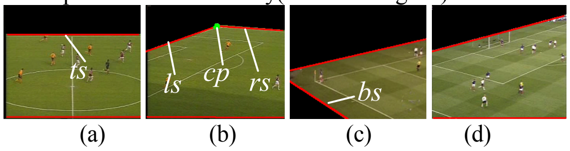


Figure 4. Shape features for classifying global views.

SVM classifiers are employed to perform the classification task on extracted features. An incremental scheme is adopt to improve the extensibility, which will be described in section 3.2. After the view classification procedure and a smooth procedure, view labels are obtained for video clips.

2.3 Descriptors extraction

For each of the soccer views, three kinds of mid-level descriptors: “view label”, “camera motion” and “shot length”, are extracted.

In a soccer video, the main camera often looks at the moving ball. When an exciting event happens, the camera often has a large movement for the large displacement of the ball.

For example, for a "shoot on goal" event, the camera may have a large movement to track the ball flying to the goalmouth. Then camera motion description is meaningful for soccer event detection. If the values of "Zoom" (z), "Pan"(p) and "Tilt"(t) for each video frame are obtained by a standard global motion estimation algorithm, we can describe the camera motion (cmd) by thresholding the values of z , p and t as

$$cmd = \begin{cases} Salient & if\ \max(z, p, t) < T_S \\ Fast\ \ Zoom & if\ \ z > T_z \\ Fast\ \ Pan & if\ \ p > T_p \\ Fast\ \ Tilt & if\ \ t > T_t \\ Low\ \ motion & otherwise \end{cases} \qquad (4)$$

where T_s is select as 0.002 which means that there are 2 pixel movement if the image size is 1000 x 1000, T_z, T_p, T_t are select as 0.02 in this paper.

On the frame labeling result, we can descript the camera motion in a labeled view by binary features as:

MS: More than 80 percent of the frames are salient	0/1
SL: Existing from salient to large pan/tilt/ zoom	0/1
LM: Existing large pan/tilt/zoom	0/1

These descriptions will contribute to the high level events to some extent. For example, *SL* will imply that a "placed kick" event. *LM* will contribute to "shot on goal" event.

Shot length is meaningful for event detection. For example, a "goal" event is often followed by several short shots to replay the event in different camera angles. We combine the color histogram, edge distribution features and corner points features for shot boundary detection by a HMM shot detect model [14]. On the obtained shot boundaries, shot length descriptor for a view is calculated as

$$sld = \begin{cases} Short\ \ shot & if\ \ L < 0.5 * L_{average} \\ Long\ \ shot & if\ \ L > 2.0 * L_{average} \\ Medium\ \ shot & otherwise\ . \end{cases} \qquad (5)$$

where L is the length of the present shot and $L_{average}$ is the average length of all played shots. By this function, we obtain shot length descriptor by discrete values in {0 (short), 1(medium), 2(long)}.

3. EXCITING EVENT DETECTION

Based on obtained the mid-level descriptors and their temporal relationship, event detection by SVM-based incremental learning is carried out in this section.

3.1 Descriptors with temporal structure

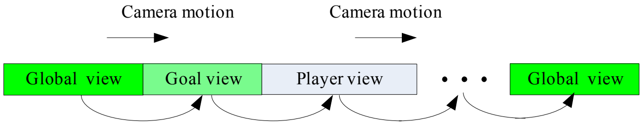
In this paper, we do not use temporal statistic modes to formulate exciting event in soccer video. The temporal relation-ship of views is captured by a fixed structure. By mid-level description, features for soccer events can be built as:

{{Descriptor of view (i)}, …, {Descriptor of view (n)} (6)

and each view is represented by

{View descriptor, motion descriptor, shot length descriptor} (7)

where n is the view index.



Temporal structure

Figure 5. Event representation

This fixed view temporal structure is more reasonable than using temporal learning models etc. HMM to learn structure event, since the latter is based on a probability framework than often need lots of training examples. While in the soccer game, exciting events for examples "shoot on goal" event is rarely happened, and then it difficult to mark large number of training samples. This is also the reason that SVM classifier is employed for event detection task.

3.2 SVM-based incremental learning

It is well know that the variance of camera position and style of TV station will produce video program of different styles. For example, "player close-up" view can vary a lot from different soccer match as shown in Fig.6. Events content and temporal structure also vary a lot in different soccer program, which makes the extensibility of event detection models quite poor

Figure 6. "Player close-up view" in different match.

An SVM-based incremental learning method is investigated to improve the extensibility of event detection model. Training an SVM "incrementally" on new data by discarding all previous data except their support vectors, gives only approximate results [15]. Supposing that a new samples is s_i and the original models contains support vector set $\{s_v\}$, when a new sample with class label is reaching, the author alternate the supports vectors of old model by keeping the KKT condition, then for the a new samples is s_i, there are four result:

1. s_i is not a new support vector
2. s_i is not a new support vector, while some old supports vectors are discarded for the arrival of s_i;
3. s_i is a new support vector,
4. s_i is a new support vector, while some old supports vectors are discarded for the arrival of s_i.

We can update the SVMs when one incremental training sample reaches without a retraining procedure. The incremental training samples can be obtained by human-computer interaction. Details of the incremental learning could be found in [15].

4. EXPERIMENTAL RESULTS

Soccer video from the Europe Cup 2004 and England soccer matches totally 15 videos are used for experiments.

For view classification, table 1 reports the experimental result. It can be seen that when training examples and test examples are extracted from the same video set, for example "Europe Cup", the classification accuracy is high. When we change the test set with "England soccer" match, the classification result will drop a lot. While with the incremental learning algorithm, the classification accuracy will increase (from 92.5% to 83.4%). When using 20-30 samples for incremental learning, the classification result near to that of without changing video set.

Table. 1 View classification result

	View classification Accuracy
Without changing video set	92.5%
Changing video set	83.4%
Incremental by 10 frames	86.7%
Incremental by 20 frames	90.1%
Incremental by 30 frames	91.0%

We selected three most representative exiting events for experiment. They are "shoot on goal", "placed kick" and "break" events. We can also found that the classification result is improved a lot with a incremental learning procedure, which can ensure that the method can be easily extended to different styles of soccer video with little human labor.

Table. 2 View classification result

	Event detection accuracy
Without changing video set	84.6%
Before incremental learning	71.1%
Incremental by 10 frames	76.4%
Incremental by 20 frames	78.4%
Incremental by 30 frames	78.9%

The event detection accuracy is about 78.9% given 72.1% recall rate after incremental learning. Although the recall and accuracy rate is not satisfying and it is even worse than some reported result on event detection, the incremental learning idea for improving extensibility is novel.

5. CONCLUSIONS

In this paper, we have presented new soccer event detection method based on mid-level description incremental learning. The incremental learning improves the extensibility of the exciting event detection method. In the future work, more effective mid-level description and systematic research on incremental learning should be carried out.

ACKNOWLEDGE

This work is supported by NEC Research China on "Context-based Multimedia Analysis and Retrieval Program" and "Science 100 Plan" of Chinese Academy of Sciences as a part of "SPISES: SPorts Video Summarization and Enrichment System" at "http://www.jdl.ac.cn/en/project/mrhomepage/En_index.htm ".

6. REFERENCE

[1] H. Pan, P. Van Beek and M.I. Sezan. "Detection of slow-motion replay segments in sports video for highlights generation," In Proc. IEEE ICASSP 2001.

[2] A. Hanjalic, "Generic approach to highlights extraction from a sport video," ICIP 2003.

[3] E.A. Murat Tekalp and A. M. Tekalp, ``Generic play-break event detection for summarization and hierarchical sports video analysis," to appear in Proc. IEEE ICME 2003.

[4] L.Y. Duan, M. Xu, T. Chua, Q. Tian, C.S Xu, "A mid-level representation framework for semantic sports video analysis", ACM Multimedia 2003.

[5] N. Babaguchi, Y.Kawai, T. Kitahashi, "Event Based Indexing of Broadcasted Sports Video by Intermodal Collaboration," IEEE Trans. On Multimedia 2002.

[6] S. Nepal, U. Srinivasan, G. Reynolds, "Automatic detection of 'Goal' segments in basketball videos," International conference on ACM Multimedia 2001.

[7] Y. Gong, T.S. Lim, and H.C. Chua, "Automatic Parsing of TV Soccer Programs", IEEE International Conference on Multimedia Computing and Systems, May, 1995

[8] L. Xie, P. Xu, S.-F. Chang, A. Divakaran and H. Sun, "Structure Analysis of Soccer Video with Domain Knowledge and Hidden Markov Models," Pattern Recognition Letters, 2004

[9] E.A. Murat Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization," IEEE Trans. on Image Process, Volume: 12, Issue: 7, July 2003

[10] R. Leonardi, P. Migliorati, and M. Prandini,"Semantic indexing of soccer audio-visual sequences: a multimodal approach based on controlled Markov chains," IEEE Trans on CSVT 2004 Vol.14, No.5.

[11] J. Assfalg, Marco Bertini,Carlo Colombo, Alberto Del Bimbo, Walter Nunziati, "Semantic annotation of soccer videos: automatic highlights identification", Computer Vision and Image Understanding, Special isssue on video retrieval and summarization, Volume 92,Issue 2/3,November/ December 2003.

[12] Y. Liu, S.Q. Jiang, Q.X. Ye, W. Gao, and Q.M. Huang, "Playfield Detection Using Adaptive GMM and Its Application," ICASSP2004.

[13] L-Y. Duan, M. Xu, T.S Chua, Q. Tian, Chang-Sheng Xu, "A mid-level representation framework for semantic sports video analysis" Proceedings of the eleventh ACM conference on Multimedia pp: 33 - 44 , 2003.

[14] Y. Liu, W. Zeng, W. Q. Wang, W. Gao, "A Novel Compressed domain Shot Segmentation Algorithm on H.264/AVC," Accepted by ICIP 2004.

[15] G. Cauwenberghs, G. and T. Poggio. "Incremental and Decremental Support Vector Machine Learning," In: Advances in Neural Information Processing Systems, MIT Press, Vol. 13, 409-415, Cambridge, MA, 2001.