Score-based Highlight Detection in Soccer Video

Bo Liu

Abstract—**With the rapid development of multimedia production and transmission technology, how to more effectively, more accurately analyze a large scale of sports video data and extract sports video summarization according to users preferences is becoming a meaningful and challenging topic. We propose a novel and computationally efficient framework for soccer video highlight detection. The proposed framework includes some new low-level soccer video processing algorithms, such as shot boundary detection, grass color region detection, view type classification, and play position classification, as well as several higher-level algorithms for slow-motion replay detection, score change detection and goal event detection. The system can output two types of summaries: i) all slow-motion replay segments in a game, ii) all goals in a game. Grass color region detection part uses EM algorithm and GMM model. View type classifier and play position classifier are based on grass region detector. The proposed framework is efficient and robust for soccer video processing. It is efficient in the sense that there is no need to compute object-based features and when cinematic features are sufficient for the detection of certain events for speedy processing. The efficiency and robustness of the proposed framework are demonstrated over a large data set, consisting of 93 soccer videos, totally more than 140 hours, captured at different in different types and countries.**

***Index Terms*—low level feature extraction, grass color region detection, score change detection, slow-motion replay detection, goal event detection, soccer video summarization**

I. Introduction

WITH the great increasing of video contents, soccer highlight detection is becoming an active research topic in multimedia analysis. There is rich literature on highlight detection []. Processing of soccer video, for example detection of slow-motion replay and goal events and creation of summaries, makes it very possible to deliver soccer video summarization that audiences are interested in over narrow band network, such as the Internet and wireless. Therefore, sports video processing needs to be completed automatically and the processing results must be semantically meaningfully. Meanwhile, manual annotation of sport videos can spend an amount of expensive costs and are usually accompanied with a strong subjectivity and a lot of annotation errors. For this situation, automatic highlight detection in soccer video can relieve the heavy annotation labor. In general, highlight in a soccer video are the special events that audiences are especially interested in, e.g. goals, shoots, fouls, corner-kicks and free kicks etc. Among all the events above, slow-motion replay and goal events weigh most. In this paper, we propose a novel soccer video processing framework that could not only satisfy these requirements but also very effectively and accurately detect replay, goal events and deliver a soccer video summarization to users.

In our framework, we think some cues in a soccer video are helpful to highlight detection. i) view type, we classify the shot view in a soccer game into major four types, including global view, medium view, close-up view and out of field view. ii), play position, we also classify play position into only two classes, e.g. middle position and side one. Additionally, as we all know, a replay part usually follows highlight event. In our system one separate image is not enough for highlight event classification, and a sequence of images can better describe the event. The framework we proposed takes soccer video analysis as the first step and output two types of summaries including slow-motion replay and goal event. Among these highlights we mentioned above, goal event detection is the focus of our topic.

In this paper, we propose a new framework for automatic, real-time soccer video analysis. The main contributions are as follows. 1) We propose new grass color region detection algorithm that are robust to variations in the domain color. 2) We introduce new algorithms for fully automatic and computation efficient detection of slow-motion replay and goal event in soccer videos. Goals are detected based solely on cinematic features resulting from the common and simple rules that score change means goal event happening. 3) Finally, we propose an efficient and effective framework for soccer video analysis and summarization that combines these algorithms in a scalable fusion.

The rest of this paper is organized as follows. We describe the novel algorithms for grass color region detection, view type classification, and play position classification in the next section. Section III presents proposed higher-level methods for slow-motion replay detection and goal event detection. Experimental results over more than 140 hours, totally 93 soccer videos from different types and countries and the temporal performance of the system are discussed in section IV.

II. Lower level feature extraction

This section explain the algorithms for lower lever feature extraction, such as grass color histogram statistics, domain color detection, view type classification and play position classification. Since view type classifier and position classifier replay on accurate detection of soccer field region in each frame, we start by presenting our robust grass color region detection.

A. Grass Color Region Detection

Grass color region detection is basic of view type and play position detector. A soccer field has one distinct dominant color (a tone of green) that vary from stadium to stadium, and also due to weather and lighting conditions within the same stadium. Therefore we do not assume any specific value for the color of the field in our system. Our only assumption is the existence of a single dominant color that indicates the soccer field. From one statistics, grass hue values in HSV color space vary from 0.15 to 0.25 on [0-1] scale from different types of soccer games.

In our system, an on-line 2-gaussion GMM model is used to describe grass color distribution. We randomly sample a certain number of frames which grass region occupies most shares from a soccer video. Here the sample rate is an experimental value, while especially in our framework the experience is that approximately 50 frames are sampled in a one-hour long soccer video.

The grass region color is only described by the hue value in each frame. In the proposed method, we firstly train a grass region color GMM model by using the Expectation-Maximization algorithm abbreviated as EM algorithm with the frames sampled from the soccer video. EM algorithm is a public one which is an iterative method for finding maximum likelihood or maximum a posteriori (MAP) estimates of parameters in statistical models, where the model depends on unobserved latent variables.

After training, we can automatically gain the means, weight and variance parameters of each Gaussian distribution of the trained GMM model. Which Gaussian distribution to choose as grass region color distribution is the next step to consider. In our system, we adopt a simple but accurate method of criterion that we firstly compute the weight/variance value of each Gaussian distribution and then select the Gaussian distribution with greater value as the grass region color distribution.

As the grass color distribution is certainly gained. Grass region colored pixels in each frame is detected by comparing each pixel with parameters of the selected Gaussian distribution. When the hue value of a pixel in a frame does not belong to the selected distribution, then the pixel is not grass colored pixel. Otherwise, if the distance of pixel hue value to the mean of distribution is larger than three time variance, the pixel also does not belong to grass colored one. Except for the two situations mentioned above, the pixels can be classified as grass region colored ones.

B. View Type classification

View type class information with other features conveys interesting semantic cues. The view type classification depends on the grass color region detection proposed above. Here our system basically adopts simple rules for view type classification. We firstly assume grass color regions are already detected accurately.

According to the proposed simple rules, we classify soccer views into the following four types: i) when grass area is greater than one threshold predefined, the image is classified as long view. ii) when player face is detected in a frame, the image is regarded as close-up view. iii) when player body is detected in a frame, it is regarded as middle view. iv) other view not belonging to any type above is classified as out-of-field. These rules above are not so complete but enough for event classification.

C. Play Position Classification

In our implementation play position classification is only evaluated in the long view. Since play position classification rely on grass region detection, grass region detection also must be put as the first step and focus. So it is assumed that the grass color region has already detected. Then we use the Connected Component Analysis algorithm in the detected grass color region. By this means, internal noise in grass color region could be removed. In addition, dilating and eroding algorithms are applied to the grass region to make region contour smooth. By component shape and dilation we can remove noise. The classification rule is defined simply as follows, if non-grass region (left and right side) rate is less than one threshold, image is regarded as middle position, otherwise is regarded as side position. Error rate is 4.5% (18/332) in long view images.

III. SCORE-BASED HIGHLIGHT DETECTION

A. Why Is Score Board

As we all know, modification of score board stand for goal event. In the score board score-related region includes time information, team name, score info and other logos. Our statistics is done in 96 videos provided by Orange Sports. Among these soccer videos, Left-Top is 92.7%, Top is 6.3% and Right-Top is nearly 1%. Where are scores in various score boards? There are only two situations. One is between team names. The other is on the right of team names. Besides, where is time info: i) on the left of team names, ii) on the bottom of team names, iii) on above of team names.

B. Score Board Detection

We take the source soccer video for input at first. In video decoding part video sample rate is 1, i.e. sampling one frame each second. Then the sampled images are converted into grayscale images because only grayscale image can generate Canny edge image. Then by using Canny edge detection algorithm, we generate Canny edge image of each frame. Next step is to calculate all the Canny edge image of each frame. In our framework, only those pixels locating in one-third top region and the abscissa between [20, width-20] (width referring to image width size) in the edge image are considered for calculating. Then we should scale the accumulation sum to [0-255]. The last step is to generate the stable edge probability image. Pixels whose value is greater than a threshold predefined are regarded as the score board and logo region. The threshold here is set as 60 in our framework.

Furthermore, in order to improve the performance of score board detection, eroding and dilating algorithms are used to detect score board region. Only the contours whose area is greater than one threshold are regarded as candidate region contours in the edge probability image generated in previous step. Here we only process the leftmost contour according to the obvious fact that the score board of almost all the game types is on the left at the default. Then the proposed system can automatically output the score board region. The error rate is about 1.04%. (only one error in 96 soccer videos).

C. Score-based Video Segmentation

In current section we assume that the score board has detected accurately and the score board region is already certain. So in the soccer video segmentation part, we should calculate matching probability of each edge image with the generated edge probability image mentioned in previous section first of all. We define matching probability as the share of the number of match pixels in edge image in the detected score board region. One thing to note is that only the pixels in the score board region are involved in calculation. Each matching probability corresponds to an edge frame. For the matching probability set, we need to delete some noises in it to make the set regular. As for the next step -- frame classification, if a value in the matching probability set is greater than one threshold pre-defined, it's corresponding frame belongs to normal playing time segment, otherwise, the frame belongs to non-playing time segment. The threshold is basically set as 0.45. The final step what we should do is to estimate the time point of segmenting soccer video. The time point when frame transferring between normal playing time and non-playing time is the starting time or ending time of a certain slow-motion replay segment and need to be saved. Because of the saved time points set is even numbered, so in the set the odd location one is the starting time of some replay segment and the even location one is the corresponding ending time.

Through the steps proposed above, our framework automatically and intelligently segment a soccer video into a plurality of video segments according to whether the score board existing or not. The segment in which score board exists continuously is normal-playing video segment, otherwise, it belongs to other video segments including slow-motion replay, advertisement, opening ceremony and so on. So we achieve the aim of score based video segment.

D. Slow-Motion Replay Detection

Replays in sports broadcast are excellent locators of semantically important segments for higher-level video processing. Several slow-motion replay detectors exist in the literature [] with complex algorithm and long computation time. However, our system proposes an efficient and robust algorithm that can detect the slow-motion replay segments in a short time and with an extremely high accurate rate. Since we only need to determine if a given segment consists of the score board region and grass color region.

The proposed algorithm is based on score-based video segmentation detector. The video segments without score board mentioned in last section are candidate segments for slow-motion replay. Given the candidate segments consist of replay, advertisement and some other video parts. The largest difference between slow-motion replay part and other video parts is that most frames in replay segment contain grass color region. According to this difference, we efficiently and robustly detect the slow-motion replay video segment. Slow-motion replay will usually follow a wonderful event, which has important significance to soccer highlight detection.

E. Goal Event Detection

Detection of certain events and objects in a soccer video game enables generation of more concise and semantically rich summaries. Since goals are arguably the most significant event in soccer, we propose a novel goal detection algorithm in current section.

The score in score region changes when there is a goal event happening. As it is known to us, there are two major regions in a score board, i.e. score region and time region. Our proposed algorithm is based on an empirical fact that the frequency of score and time regions changes more quickly in time than other regions such as team name and borders. Therefore, we come out a fast and efficient detection algorithm for score and time region detection. The steps of process are as follows. Firstly, we remove those edge images of frames that are not in normal playing time video segments to update the set of Canny edge images. Secondly, in each normal-playing video segment, we generate its own sub edge probability images, in which specific algorithm is same as generation of edge probability of entire soccer video. Thirdly, we compare each sub probability with the other ones, the purpose of doing that is to strengthen difference values of pixels in score and time regions of each sub edge probability image. Fourthly, calculate difference value of each pixel, specific process is that calculate difference value of corresponding pixel in any two sub edge probability image. If difference value is greater than one threshold pre-defined, the pixel is regarded as score and time region, otherwise, it belongs to other region. Finally, scale the generated probability grayscale image to [0-255].

Goal event is considered as the most excellent one in a soccer video. In goal detection section, we only compare the edge probability image generated in each normal-playing video segment with the adjacent one. There the difference from the score and time region detection algorithm is that in the score and time region detection algorithm one sub edge probability image are compared with all the other ones. Then we accumulate the difference value of all the pixels, and normalize the cumulative difference value to [0-255] scale.

If the normalized cumulative difference value is greater than a threshold (set at 50), which indicates that there are mutations happened in score or time region in current normal-playing video segment. It should be noted that the mutation does not only change in the score region, time region may also happen to change. However, according to an empirical fact that the frequency of time region changes in timeline is much faster than the score region, in addition, the number of mutations in time region is much more. Therefore, time region can be removed on the basis of such an experience above, only leaving the changes in score region we need. Therefore score change in score board indicates that a goal event happens in current video segment. Meanwhile the goal event is accurately detected and goal time is estimated.

IV. Remaining Problems

Although the novel system we proposed could quickly and efficiently achieve our aim of replay detection and goal event detection and so on, there are several remaining problems that still need to be improved with effective and complete algorithms.

Grass color region detection is basic of view type classification and play position classification. However, color of the grass field region may vary from stadium to stadium and sometimes could be affected easily by the weather. So grass region color detection may be not robust. Furthermore, audience background may affect the result of domain color detection a lot in grass region detection part. Finally, the number of GMM model Gaussians could not be definite, but it ranges from 2 to 4 according to our experiments, mostly 2-gaussians.

Additionally, in the score board detection part, there are also some problems remained. One common situation is that score board may be transparent or semitransparent in some types of soccer videos. Another situation may sometimes happen that some information about the soccer game may appear in score board, such as red card, injury time and so on.

V. Results

We have largely tested the proposed algorithms on a data set of more than 140 hours, totally 93 soccer videos. These sports videos this paper trained and tested are mainly provided by Orange Sports as a research material. Table I shows the whole game types and the number of each type.

A. Results for Low-Level Algorithms

None.

B. Results for Higher-Level Video Analysis

None.

C. Results for Slow-Motion Detection

None.

D. Results for Goal Event Detection

Goals are detected in 93 test soccer videos in the database.

VI. Conclusion

In this paper, a new framework for summarization of soccer video has been introduced. The implication of the proposed framework include grass color region detection, efficient and robust slow-motion replay detection, goal event detection and summarization according user preferences ,each of which makes the system highly desirable. The topic for future work include 1) integration of aural and textual features to increase the accuracy of event detection and 2) extension of the proposed framework to different sports, such as rugby and tennis, which require different event and object detection modules.

References

Basic format for books:

[1] J. K. Author, "Title of chapter in the book," in *Title of His Published Book, x*th ed. City of Publisher, Country if not

[2] USA: Abbrev. of Publisher, year, ch. *x*, sec. *x*, pp. *xxx–xxx.*

Examples:

[3] G. O. Young, "Synthetic structure of industrial plastics," in *Plastics,* 2nd ed., vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.

[4] W.-K. Chen, *Linear Networks and Systems*. Belmont, CA: Wadsworth, 1993, pp. 123–135.

Basic format for periodicals:

[5] J. K. Author, "Name of paper," *Abbrev. Title of Periodical*, vol. *x,* no. *x*, pp. *xxx-xxx*, Abbrev. Month, year.

Examples:

[6] J. U. Duncombe, "Infrared navigation—Part I: An assessment of feasibility," *IEEE Trans. Electron Devices*, vol. ED-11, no. 1, pp. 34–39, Jan. 1959.

[7] E. P. Wigner, "Theory of traveling-wave optical laser," *Phys. Rev*., vol. 134, pp. A635–A646, Dec. 1965.

[8] E. H. Miller, "A note on reflector arrays," *IEEE Trans. Antennas Propagat*., to be published.

Basic format for reports:

[9] J. K. Author, "Title of report," Abbrev. Name of Co., City of Co., Abbrev. State, Rep. *xxx*, year.

Examples:

[10] E. E. Reber, R. L. Michell, and C. J. Carter, "Oxygen absorption in the earth's atmosphere," Aerospace Corp., Los Angeles, CA, Tech. Rep. TR-0200 (4230-46)-3, Nov. 1988.

[11] J. H. Davis and J. R. Cogdell, "Calibration program for the 16-foot antenna," Elect. Eng. Res. Lab., Univ. Texas, Austin, Tech. Memo. NGL-006-69-3, Nov. 15, 1987.

Basic format for handbooks:

[12] Name of Manual/Handbook, *x* ed., Abbrev. Name of Co., City of Co., Abbrev. State, year, pp. *xxx-xxx.*

Examples:

[13] *Transmission Systems for Communications*, 3rd ed., Western Electric Co., Winston-Salem, NC, 1985, pp. 44–60.

[14] *Motorola Semiconductor Data Manual*, Motorola Semiconductor Products Inc., Phoenix, AZ, 1989.

Basic format for books (when available online):

[15] Author. (year, month day). *Title.* (edition) [Type of medium]. *volume (issue).* Available: site/path/file

Example:

[16] J. Jones. (1991, May 10). *Networks.* (2nd ed.) [Online]. Available: http://www.atm.com

Basic format for journals (when available online):

[17] Author. (year, month). Title. *Journal.* [Type of medium]. *volume (issue),* pages. Available: site/path/file

Example:

[18] R. J. Vidmar. (1992, Aug.). On the use of atmospheric plasmas as electromagnetic reflectors. *IEEE Trans. Plasma Sci.* [Online]. *21(3),* pp. 876–880. Available: http://www.halcyon.com/pub/journals/21ps03-vidmar

Basic format for papers presented at conferences (when available online):

[19] Author. (year, month). Title. Presented at Conference title. [Type of Medium]. Available: site/path/file

Example:

[20] PROCESS Corp., MA. Intranets: Internet technologies deployed behind the firewall for corporate productivity. Presented at INET96 Annual Meeting. [Online]. Available: http://home.process.com/Intranets/wp2.htp

Basic format for reports and handbooks (when available online):

[21] Author. (year, month). Title. Company. City, State or Country. [Type of Medium].Available: site/path/file

Example:

[22] S. L. Talleen. (1996, Apr.). The Intranet Archi-tecture: Managing information in the new paradigm. Amdahl Corp., CA. [Online]. Available: http://www.amdahl.com/doc/products/bsg/intra/infra/html

Basic format for computer programs and electronic documents (when available online): ISO recommends that capitalization follow the accepted practice for the language or script in which the information is given.

Example:

[23] A. Harriman. (1993, June). Compendium of genealogical software. *Humanist*. [Online]. Available e-mail: HUMANIST@NYVM.ORG Message: get GENEALOGY REPORT

Basic format for patents (when available online):

[24] Name of the invention, by inventor's name. (year, month day). *Patent Number* [Type of medium]. Available: site/path/file

Example:

[25] Musical toothbrush with adjustable neck and mirror, by L.M.R. Brooks. (1992, May 19). *Patent D 326 189* [Online]. Available: NEXIS Library: LEXPAT File: DESIGN

Basic format for conference proceedings (published):

[26] J. K. Author, "Title of paper," in *Abbreviated Name of Conf.*, City of Conf., Abbrev. State (if given), year, pp. *xxxxxx.*

Example:

[27] D. B. Payne and J. R. Stern, "Wavelength-switched pas- sively coupled single-mode optical network," in *Proc. IOOC-ECOC,* 1985, pp. 585–590.

Example for papers presented at conferences (unpublished):

[28] D. Ebehard and E. Voges, "Digital single sideband detection for interferometric sensors," presented at the 2nd Int. Conf. Optical Fiber Sensors, Stuttgart, Germany, Jan. 2-5, 1984.

Basic format for patents:

[29] J. K. Author, "Title of patent," U.S. Patent *x xxx xxx*, Abbrev. Month, day, year.

Example:

[30] G. Brandli and M. Dick, "Alternating current fed power supply," U.S. Patent 4 084 217, Nov. 4, 1978.

Basic format for theses (M.S.) and dissertations (Ph.D.):

[31] J. K. Author, "Title of thesis," M.S. thesis, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year.
[32] J. K. Author, "Title of dissertation," Ph.D. dissertation, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year.

Examples:

[33] J. O. Williams, "Narrow-band analyzer," Ph.D. dissertation, Dept. Elect. Eng., Harvard Univ., Cambridge, MA, 1993.
[34] N. Kawasaki, "Parametric study of thermal and chemical nonequilibrium nozzle flow," M.S. thesis, Dept. Electron. Eng., Osaka Univ., Osaka, Japan, 1993.

Basic format for the most common types of unpublished references:

[35] J. K. Author, private communication, Abbrev. Month, year.
[36] J. K. Author, "Title of paper," unpublished.
[37] J. K. Author, "Title of paper," to be published.

Examples:

[38] A. Harrison, private communication, May 1995.
[39] B. Smith, "An approach to graphs of linear forms," unpublished.
[40] A. Brahms, "Representation error for real numbers in binary computer arithmetic," IEEE Computer Group Repository, Paper R-67-85.

Basic format for standards:

[41] *Title of Standard*, Standard number, date.

Examples:

[42] IEEE Criteria for Class IE Electric Systems, IEEE Standard 308, 1969.
[43] Letter Symbols for Quantities, ANSI Standard Y10.5-1968.