

SVM-based Soccer Video Summarization System

Hossam M. Zawbaa
Cairo University,
Faculty of Computers and Information
Cairo, Egypt
Email: hossam.zawba3a@gmail.com

Nashwa El-Bendary
Arab Academy for Science,
Technology, and Maritime Transport
Cairo, Egypt
Email: nashwa_m@aast.edu

About Ella Hassanien
Cairo University,
Faculty of Computers and Information
Cairo, Egypt
Email: aboitcairo@gmail.com

Ajith Abraham

IT for Innovations, EU Center of Excellence, Faculty of Electrical Engineering and Computer Science
VSB - Technical University of Ostrava, Ostrava - Poruba, Czech Republic
Machine Intelligence Research Labs (MIR Labs)
Scientific Network for Innovation and Research Excellence, WA, USA
Email: ajith.abraham@ieee.org

Abstract—In this paper, we propose a system for soccer video summarization using support vector machine (SVM). The proposed system initially segments the whole video stream into small video shots. Then, the system applies support vector machine (SVM) algorithm for emphasizing important segments with logo appearance with addition to detecting the caption region providing information about the score of the game. Subsequently, the system uses k-means algorithm and Hough line transform for detecting vertical goal posts and Gabor filter for detecting goal net. Finally the system highlights the most important events during the match. Experiments on real soccer videos demonstrate encouraging results. The proposed system greatly reduces workload and enhances the accuracy of summarizing soccer video matches with reference to both recall and precision performance measurement criteria.

Keywords-SVM; Video summarization; Soccer video processing;

I. INTRODUCTION

Soccer video matches always attract major sports audience. Recently, the amount of digitized video content has been increasing rapidly and users need to access their content through various network facilities. As watching a soccer match needs a lot of time, many TV fans of sport competitions prefer to watch a summary of football games [1]. According to this, soccer video analysis has recently attracted much research and a wide spectrum of possible applications have been considered. Traditionally soccer videos were analyzed manually but it costs valuable time. Therefore it is necessary to have a tool that does the job automatically.

The support vector machine (SVM) algorithm for automatic logo detection and score board detection and from broadcast soccer videos. the support vector machine (SVM) is a theoretically superior machine learning (ML) methodology with great results in classification of high dimensional datasets and has been found competitive with the best ML algorithms [2].

This paper presents a soccer video summarization system using support vector machine (SVM). The proposed system is composed of six phases; namely, pre-processing phase, shot processing phase, SVM-based logo replay detection phase, SVM-based score board detection phase, excitement event detection phase, and finally logo-based event detection and summarization phase. The rest of this paper is organized as follows. Section II gives an overview of SVM machine learning technique. Section III presents the different phases of the proposed automatic soccer video summarization system. Section IV shows the obtained experimental results. Finally, Section V addresses conclusions and discusses future work.

II. SUPPORT VECTOR MACHINE (SVM): A BRIEF BACKGROUND

The support vector machine (SVM) algorithm seeks to maximize the margin around a hyperplane that separates a positive class from a negative class.

Given a training dataset with n samples $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, where x_i is a feature vector in a v -dimensional feature space and with labels $y_i \in -1, 1$ belonging to either of two linearly separable classes C_1 and C_2 . Geometrically, the SVM modeling algorithm finds an optimal hyperplane with the maximal margin to separate two classes, which requires to solve the optimization problem, as shown in equations (1) and (2).

$$\text{maximize} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j \cdot K(x_i, x_j) \quad (1)$$

$$\text{Subject - to : } \sum_{i=1}^n \alpha_i y_i, 0 \leq \alpha_i \leq C \quad (2)$$

where, α_i is the weight assigned to the training sample x_i . If $\alpha_i > 0$, x_i is called a support vector. C is a

regulation parameter used to trade-off the training accuracy and the model complexity so that a superior generalization capability can be achieved. K is a kernel function, which is used to measure the similarity between two samples. Different choices of kernel functions have been proposed and extensively used in the past and the most popular are the gaussian radial basis function (RBF), polynomial of a given degree, and multi layer perceptron. These kernels are in general used, independently of the problem, for both discrete and continuous data.

III. THE PROPOSED SVM-BASED SOCCER VIDEO SUMMARIZATION SYSTEM

The SVM-based soccer video summarization system proposed in this paper is composed of six fundamental building phases; 1) pre-processing phase that segments the whole video stream into small video shots, 2) shot processing phase that applies two types of classification to the video shots resulted from the pre-processing phase, 3) replay detection phase that applies support vector machine (SVM) algorithm for emphasizing important segments with logo appearance, 4) score board detection phase that uses SVM algorithm for detecting the caption region providing information about the score of the game, 5) excitement event detection phase that detects both vertical goal posts and goal net using k-means algorithm and Hough line transform for detecting goal posts Gabor filter for detecting goal net, and finally 6) logo-based event detection and summarization phase that highlights the most important events during the match. These six phases are described in detail in this section along with the steps involved and the characteristics feature for each phase.

A. Pre-processing Phase

The goal of this phase is to segment the whole video stream into small video shots. Firstly, the dominant color in the video frame is detected, then the shot boundary detection algorithm in [3] is applied in order to output video shots based on dominant color derived features [4].

1) *Grass dominant color extraction*: The dominant color is the color that fills most of the given area, and it is different among various play-fields. In this paper, we are concerning only with the soccer game, which has a green color for the playing field. As dominant color extraction is challenging due to effects on the play-field such as shadow, lighting, low resolution and other environmental factors. [4].

2) *Shot boundary detection*: The separated views that come from multiple cameras positioned at different locations. It can be realized that while changing from one camera to another, this indicates a start of a new shot and marks a boundary of a new shot. [3]–[5]. Algorithm (1) shows the steps of shot boundary detection.

Algorithm 1 Shot Boundary Detection

```

1: for each frame do
2:   Convert the frame from RGB to HSI color space
3:   Do frame skipping by step ( $k = 10$  frames) to convert
   a gradual transition into a cut transition
4:   Calculate the motion difference= the difference be-
   tween current hue frame and the next hue k-frame
5:   Divide the original color frame into 32 x 32 blocks
   size
6:   Compute the mean change = the average blocks
   change between the current frame and the next k-
   frame
7:   Compute the total percentage of changed blocks =
   the total percentage of changing blocks between the
   current frame and the next k-frame
8:   Compute the grass ratio = the average difference of
   grass dominate color between the current frame and
   the next k-frame
9:   if (Mean change > Thr) and (total percentage of
   changing blocks > 0.25) and (grass ratio > 0.1) then
10:     Mark a new shot
11:   end if
12: end for

```

B. Shot Processing Phase

This phase applies two types of classification; namely, shot-type classification and play/break classification, to the video shots resulted from the pre-processing phase.

1) *Shot-type classification*: Different shot-types can be used in order to make different scenes that can be used for high-level video analysis. Cinematographers classify a shot into one of four categories; namely, long, medium, close-up, and audience (out-of-field) shot classes. A set of thresholds have been defined for distinguishing the grass-ratio for the different shot-types [3]. For the proposed system, we applied four threshold ratios, each frame can be classified into one of the previously stated views [4].

2) *Play/break classification*: Using the start and end frames location of view-type classified shots, the boundaries of each play and break shot can be determined. Consecutive play shots are considered as a play scene, which usually are ended with a consecutive break shots. Thus, a play-break sequence is a combination of consecutive play and break scenes, and sport games consist of many of this sequences [4], [6].

C. Replay Detection Phase

Replay is a video editing technique that is often used to emphasize an important segment with a logo appearance for one or several times. In sports video, there is often a highlighted logo that appears at the start and end of a replay segment, which indicates an exciting event within the soccer



Figure 1. The score board caption region



Figure 2. Hough transform detection for the vertical goal posts

match [4], [7]. Algorithm (2) describes the steps of Logo detection algorithm.

Algorithm 2 Logo Detection Using SVM Classifier

- 1: Train the SVM classifier with correct logo and false logo samples
 - 2: **for** Each frame **do**
 - 3: Adjust image intensity values for increasing the contrast of the input frame
 - 4: Select region of interest based on color for returning a binary image
 - 5: Calculate frame white ratio = the percentage of white pixels in the whole frame
 - 6: **if** Any frame contains a large contrast object (the white frame ratio be greater than 0.5) **then**
 - 7: Get the original colored frame for the classification
 - 8: **if** The logo is real **then**
 - 9: Mark this shot as replay shot
 - 10: **end if**
 - 11: **end if**
 - 12: **end for**
-

D. Score Board Detection Phase

The score board is a caption region distinguished from the surrounding region, which provides information about the score of the game or the status of the players [8]. The caption often appears at the bottom part of image frame for a short while and then disappears almost after appearing for 5 seconds. When the score board is detected with enough confidence, it can undoubtedly provide the inference of goal event, because after every scored goal the score board is displayed. Figure 1 shows different types of the existence of the score board caption region. The lower third of each frame was checked for containing a score board by applying algorithm (2).

E. Excitement Event Detection Phase

Most exciting events occur in the goal-mouth area such as goals, shooting, penalties, direct free kicks, etc. Other non-exciting events such as dull passes in the mid-field, defense and offense or some other shots to the audiences or coaches, are not considered as exciting as the former

events [9]. Excitement event detection is based on three features; namely, 1) vertical goal posts detection, 2) goal net detection, and 3) audio loudness detection.

1) *Vertical goal posts detection:* The two vertical goal posts are distinctively characterized by their vertical strips of white and grow connected pixel gray values of white. Hough transform is used for detecting the two goal posts, as shown in figure 2. Algorithm (3) presents the steps applied to each frame for detecting the vertical goal posts.

Algorithm 3 Vertical Goal Posts Detection

- 1: **for** each frame **do**
- 2: Use K-means clustering to convert each frame to binary image using squared Euclidean distances measure
- 3: Given a set of observations $(x_1, x_2, , x_n)$, where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k sets ($k \leq n$) $S = S_1, S_2, , S_k$ so as to minimize the within-cluster sum of squares (WCSS)

$$arg\,smin \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2 \quad (3)$$

where, μ_i is the mean of points in S_{i-1}

- 4: Use Hough transform to detect the two goal posts

$$rho = x * cos(\theta) + y * sin(\theta) \quad (4)$$

where, rho is the distance from the origin to the line along a vector perpendicular to the line, and θ is the angle between the x-axis and this vector

- 5: **if** The overlap between the vertical parallel lines greater than 80% **then**
 - 6: mark this frame as goal post frame
 - 7: **end if**
 - 8: **end for**
-

2) *Goal net detection:* Detection of the two vertical goal posts isn't sufficient for possible exciting play. So, there still a need for an extra step to increase the accuracy of goal-mouth appearances detection. Accordingly, the proposed system checks goal post frames for goal net existence using Gabor filter [10]. The Gabor filter is used due to that the

goal net has a unique pattern and repeated many times.

The Gabor filter is basically a Gaussian filter, with variances s_x and s_y along x and y -axes, respectively. the s_x and s_y are modulated by a complex sinusoid, with center frequencies U and V along x and y -axes, respectively. The Gabor filter is described by equations (5), (6), and (7).

$$G = \exp\left(\left(-\frac{1}{2}\left(\frac{\hat{x}}{s_x}\right)^2 + \left(\frac{\hat{y}}{s_y}\right)^2\right)\right) * \cos 2 * \Pi * f * \hat{x} \quad (5)$$

$$\hat{x} = x * \cos(\theta) + y * \sin(\theta); \quad (6)$$

$$\hat{y} = y * \cos(\theta) - x * \sin(\theta); \quad (7)$$

Where, s_x and s_y : variances along x and y -axes, respectively, f : frequency of the sinusoidal function, θ : the orientation of Gabor filter, and G : The output filter.

F. Audio Loudness Detection

Loudness, silence and pitch generated by a commentator and/or crowd are effective measurements for detecting excitement. The volume of each audio frame is calculated using equation (8):

$$Volume = \frac{1}{N} * \sum_{n=1}^N |x(n)| \quad (8)$$

Where N is the number of frames in a clip and $x(n)$ is the sample value of the n th frame. To calculate pitch and silence, we applied the sub-harmonic-toharmonic ratio based pitch determination in [11] for its reliability. Louder, less silence, and higher pitch audio frames are identified by using dynamic thresholds presented in [12]. So, we can detect the excitement shots.

G. Event Detection and Summarization Phase

The summarized segment may contain only important events, such as: goal shots, attacks, or penalty shots [3]. The proposed system highlights the most important events during the soccer match, such as goals and goal attempts, in order to save the viewer's time and introduce the technology of computer-based summarization into sports field. figure 3 shows the different event type classification.

1) *Goal event detection*: A goal is scored when the whole soccer ball passes the goal line between the goal posts and under the crossbar. However, it is difficult to verify these conditions automatically and reliably by the state-of-the-art video processing algorithms. The occurrence of a goal event leads to a break in the game [13]. Figure 4 illustrates the sequence of cinematic features after scoring a goal. Finally, the restart of the game is usually captured by a long shot.

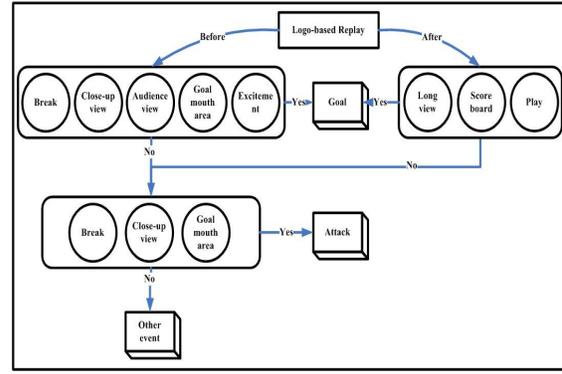


Figure 3. Event type classification

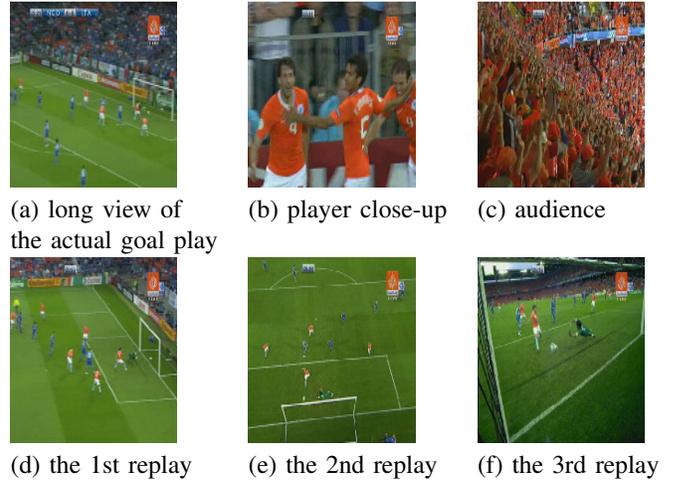


Figure 4. An example of goal broadcast: the temporal order is from (a) to (f)

2) *Attack and other event detection*: Attack events may also match a lot of goal event features, although not as consistently as goals. The addition of attack events in the summaries may even be desirable since each of these events consists of interesting shots [3]. There are other interesting events such as: fouls, cards, injure, or offside. The addition of these events in the summaries may even be desirable in order for each event to contain of interesting shots. Therefore, more of users may enjoy watching interesting fouls and offside events.

IV. EXPERIMENTAL RESULTS

The proposed system was evaluated using five videos for soccer matches from: World Cup Championship 2010, Africa Championship League 2010, Africa Championship League 2008, European Championship League 2008, and Euro 2008. All soccer videos are in Audio Video Interleave (AVI) format with a frame rate of 30 fps and an audio track

Table I
EVALUATION OF LOGO BASED REPLAY USING SVM

Factors	SVM
Duration (hh:mm:ss)	1:53:39
Correct	103
False	8
Miss	2
Recall	98.1%
Precision	92.8%

Table II
EVALUATION OF SCORE BOARD DETECTION

Duration (hh:mm:ss)	Correct	False	Miss	Recall	Precision
1:53:39	68	5	1	98.5%	93.1 %

that is sampled at 44.1 kHz.

Table I illustrates results of logo based replay detection stage. Attained recall and precision ratios for logo based replay detection are 98.1% and 92.8%, respectively.

Table II and table III show the results of score board and goal mouth detection, respectively. For score board detection, SVM classifier has been used whereas both Gabor filter and Hough transform have been used for goal mouth detection.

Table IV shows the confusion matrix for event detection and summarization resulted from the proposed system.

V. CONCLUSIONS AND FUTURE WORKS

The SVM-based system proposed in this paper for broadcast soccer videos summarization was evaluated using videos for soccer matches of five international soccer championships. The proposed system is composed of six phases; namely, pre-processing phase, shot processing phase, replay detection phase, score board detection phase, excitement event detection phase, and logo-based event detection and summarization phase. The proposed system performs very well as its analysis results achieve high accuracy. Experiments show that the system has attained very high precision and reasonable recall ratios. For future research, we can increase the number of soccer videos and championships being examined in order to get more accurate results. Moreover,

Table III
EVALUATION OF GOAL MOUTH DETECTION

Duration (hh:mm:ss)	Correct	False	Miss	Recall	Precision
1:30:42	247	25	11	95.7%	90.8%

Table IV
CONFUSION MATRIX FOR EVENT DETECTION AND SUMMARIZATION

Event Detection	Goal	Attack	Other events
Goal	57	3	0
Attack	6	176	8
Other events	0	18	283
Recall	95%	92.6%	94%
Precision	90.5%	89%	97.3%

different machine learning techniques may be applied such as neural networks.

REFERENCES

- [1] Lotfi, E. and Pourreza, H.R.: Event Detection and Automatic Summarization in Soccer Video. 4th Iranian Conference on Machine Vision and Image Processing (MVIPO7), Mashhad, Iran, (2007).
- [2] Franc, V., Zien, A., and Schlkopf, B.: Support vector machines as probabilistic models. In Proceedings of the 28th International Conference on Machine Learning (ICML'11), ACM, pp. 665-672, (2011).
- [3] Ekin, A.: Sports Video Processing for Description, Summarization and Search. PhD Thesis, University of Rochester, Rochester, (2003).
- [4] Zawbaa, H.M., El-Bendary, N., Hassanien, A.E., Schaefer, G., and Yeo, S.S.: Logo Detection in Broadcast Soccer Videos Using Support Vector Machine, Submitted to: The 2011 Online Conference On Soft Computing in Industrial Applications WWW, (WSC16), (2011).
- [5] Xing-hua, S. and Jing-yu, Y.: Inference and retrieval of soccer event. Journal of Communication and Computer, vol. 4(3), (2007).
- [6] Tjondronegoro, D., Chen, Y.P., and Pham, B.: The power of play-break for automatic detection and browsing of self-consumable sport video highlights. In Multimedia Information Retrieval, pp. 267-274, (2004).
- [7] Ren, R. and Jose, J.M.: Football video segmentation based on video production strategy. In: The 27th European Conference on IR Research (ECIR 2005), Santiago de Compostela, Spain, pp. 433-446, (2005).
- [8] Huang, C.-L., Shih, H.-C., and Chao, C.-Y.: Semantic analysis of soccer video using dynamic Bayesian network. IEEE Transactions on Multimedia, Vol. 8, No. 4, (2006).
- [9] Zhao, Z., Jiang, S., Huang, Q., and Ye, Q.: Highlight summarization in soccer video based on goalmouth detection. Asia-Pacific Workshop on Visual Information Processing, (2006).
- [10] WAN, K., YAN, X., YU, X., and XU, C.: Real-time Goal-Mouth Detection in MPEG Soccer Video. In: Proceedings of ACM MM 2003, Berkeley, USA, pp. 311314, (2003).
- [11] Sun, X.: Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio. In: The IEEE International Conference on Acoustics, Speech, Signal Processing (ICASSP'02), Orlando, Florida, USA, Vol. 1, pp 333-336, (2002).
- [12] Tjondronegoro, D., Chen, Y.P., and Pham, B.: Sports video summarization using highlights and play-breaks. In The fifth ACM SIGMM International Workshop on Multimedia Information Retrieval (ACM MIR'03), Berkeley, USA, pp. 201-208, (2003).
- [13] Ekin, A., Tekalp, A.M., and Mehrotra, R.: Automatic Soccer Video Analysis and Summarization. IEEE Transactions on Image processing, Vol. 12, No. 7, (2003).