Event Detection Based Approach for Soccer Video Summarization Using Machine learning

Hossam M. Zawbaa[1], Nashwa El-Bendary[2], Aboul Ella Hassanien[1], and Tai-hoon Kim[3]
[1]Cairo University, Faculty of Computers and Information, Cairo, Egypt
Email: hossam.zawba3a@gmail.com, aboitcairo@gmail.com
[2]Arab Academy for Science, Technology, and Maritime Transport, Cairo, Egypt
Email: nashwa_m@aast.edu
[3]Hannam University, Korea
Email: taihoonn@hannam.ac.kr

Abstract

Many soccer fans prefer to watch a summary of football games as watching a whole soccer match needs a lot of time. Traditionally, soccer videos were analyzed manually, however this costs valuable time. Therefore, it is necessary to have a tool for doing the video analysis and summarization job automatically. Automatic soccer video summarization is about extracting important events from soccer matches in order to produce general summaries for the most important moments in which soccer viewers may be interested. This paper presents a machine learning (ML) based event detection and summarization system for emphasizing important events during soccer matches. The proposed system firstly segments the whole video stream into small video shots, then it classifies the resulted shots into different shot-type classes. Afterwards, the system applies two machine learning algorithms, namely; support vector machine (SVM) and artificial neural network (ANN), 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 detects vertical goal posts and goal net. Finally, the most important events during the match are highlighted in the resulted soccer video summary. Experiments on real soccer videos demonstrate encouraging results. The proposed approach greatly reduces workload and enhances the accuracy of summarizing soccer video matches with reference to both recall and precision performance measurement criteria.

Keywords: support vector machine (SVM), artificial neural network (ANN), machine learning (ML), soccer video summarization, hough transform, logo-based detection, replay detection, soccer event detection

1 Background and Related Work

Soccer video analysis is concerned with the extraction of valuable semantics by efficient and effective processing of combination of visual, audio and text information. However, one of the major limitations of current soccer analysis is the semantic gap between the low-level features such as (color, texture, shape and motion) and high-level representation such as (shot types, shot length and shot replays).

Most sports games are naturally organized into successive and alternating plays of offence and defence, cumulating at events such as goal or attack. If a sport video can be segmented according to these semantically meaningful events, it then can be used in numerous applications to enhance their values and enrich the user's viewing experiences [1].

Soccer is one of the most popular team sports all over the world due to the relative simplicity of its rules and the small amount of required equipment [2]. As watching a soccer match needs a lot of time, many TV fans of sport competitions prefer to watch a summary of football games [3]. 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.

Analyzing general sport games is still an open problem because of the variance and diversity of different games. Some former researchers have proposed many highlight summarization methods both for general sports game and for a specific kind of sports game. In [4], authors proposed a system for detecting the play and break event in sports videos to generate the summary. Some other researchers summarize sports videos using slow motion replays, as in [5,6]. In [7], authors proposed a goal detection method using SVM classifiers in soccer videos.

On the other hand, another group of researchers turn to study specific sports game such as basketball. For example, in [8,9] investigated the event detection method for basketball videos. [9] researched on the basketball game and proposed the representation of the goal event in basketball game. Moreover, in [10], authors investigated the highlight summarization method in racquet sports such as tennis and table-tennis.

In addition, certain features may be used to recognize the major events. Some of these features are slow motion, spectators' excitement, subtitles or other types of texts on the screen, etc. These features are extracted from sound and video sources and are called "cinematic features". Some researchers have only used these features for summarizing soccer matches, such as [11] that only uses sound to generate the summarized version, and [12] that uses the camera motion parameters to detect the major events of a match. Authors of [13] used object-base features for recognizing major events, however in [14] object trajectories and relations were used to do so.

Recently, some work on replay detection has been achieved [15]. In [5], a zero-crossing measure for both the frequency and amplitude of the fluctuations of adjacent frame difference was considered. Also, a single field inside a replay must be pinpointed in advance. Later, authors of [6] proposed another replay detection approach based on replay-logo. Firstly, they used the method, proposed in [5], to detect two replay segments. Then, they searched two frames most similar at one of the two search areas preceding the two replay segments. A verification procedure was employed at last. In [16, 17], replay detection approach for sports/non-sports video classification was introduced. Their methods were block-based, which used motion vectors in the MPEG-1 domain.

This paper presents an approach for automatic soccer videos summarization using machine learning techniques. In order to generate effective summaries for soccer videos, 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. The rest of this paper is organized as follows. Section 2 gives an overview of SVM and ANN machine learning techniques. Section 3 presents the different phases of the proposed automatic soccer video summarization system. Section 4 shows the obtained experimental results. Finally, Section 5 addresses conclusions and discusses future work.

2 Machine Learning (ML): A Brief Background

2.1 Artificial Neural Network (ANN)

Artificial neural networks (ANN) or simply neural networks (NN) have been developed as generalizations of mathematical models of biological nervous systems. In a simplified mathematical model of the neuron, the effects of the synapses are represented by connection weights that modulate the effect of the associated input signals, and the nonlinear characteristic exhibited by neurons is represented by a transfer function. There are a range of transfer functions developed to process the weighted and biased inputs, among which four basic transfer functions widely adopted for multimedia processing [18].

The neuron impulse is then computed as the weighted sum of the input signals, transformed by the transfer function. The learning capability of an artificial neuron is achieved by adjusting the weights in accordance to the chosen learning algorithm. The behavior of the neural network depends largely on the interaction between the different neurons. The basic architecture consists of three types of neuron layers: input, hidden and output layers.

In feed-forward neural networks, the signal flow is from input to output units strictly in a feed-forward direction. The data processing can extend over multiple units, but no feedback connections are present, that is, connections extending from outputs of units to inputs in the same layer or previous layers. There are several other neural network architectures (Elman network, adaptive resonance theory maps, competitive networks, etc.) depending on the properties and requirement of the application [19].

2.2 Support Vector Machine (SVM)

The support vector machine (SVM) algorithm seeks to maximize the margin around a hyperplane that separates a positive class from a negative class [20]. Given a training dataset with n samples $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, 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).

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

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

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.

3 The Proposed Soccer Video Summarization Approach

The machine learning based soccer video summarization system proposed in this paper is composed of five fundamental phases as follows:

- *Pre-processing phase* that segments the whole video stream into small video shots.
- *Shot processing phase* that applies two types of classification to the video shots resulted from the pre-processing phase.
- *Replay detection phase* that applies support vector machine (SVM) and artificial neural network (ANN) algorithms for emphasizing important segments with logo appearance.
- *Excitement event detection phase* that uses both machine learning algorithms for detecting the scoreboard which contain an information about the score of the game. And the proposed system also 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.
- *Event detection and summarization phase* that highlights the most important events during the match.

Figure 1 depicts the building phases of the proposed system. These phases are described in detail in this section along with the steps involved and the characteristics feature for each phase.

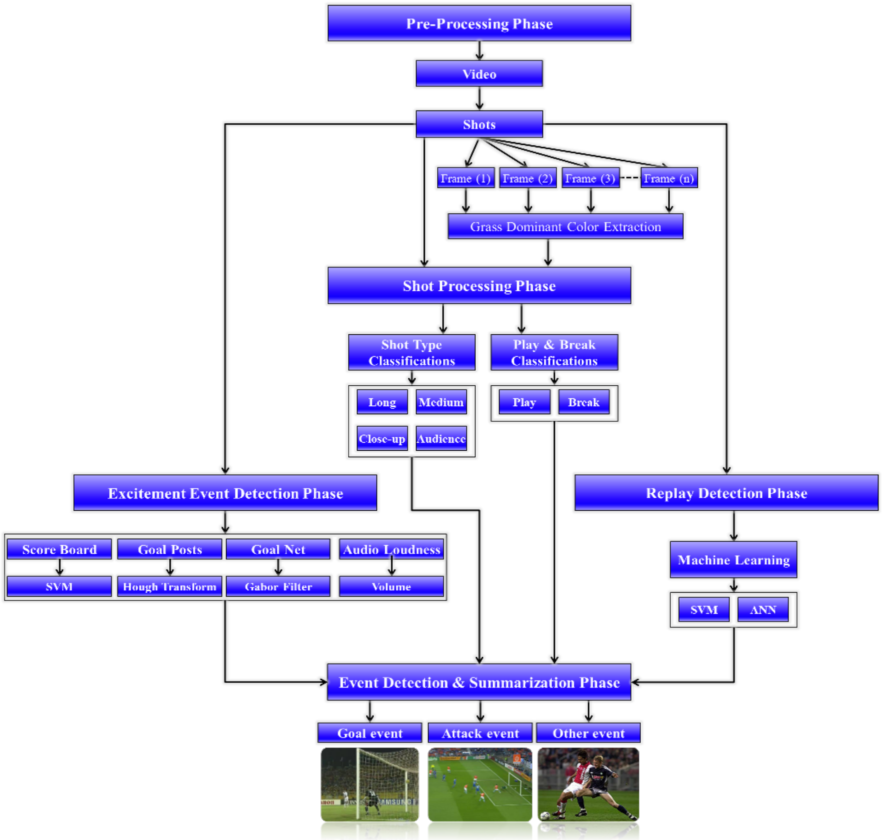

Figure 1. General system architecture

3.1 Pre-processing Phase

The goal of this phase is to segment the whole video stream into small video shots. By firstly detecting the dominant color in the video frame, then a shot boundary detection algorithm is applied in order to output video shots based on dominant color derived features [21–23].

3.1.1 Grass dominant color extraction

The dominant color is the color that fills most of the given area, and it is different for 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, there are several color spaces that have been used for the dominant color detection including HSI and RGB. Algorithm (1) shows the steps for grass dominant color extraction.

Algorithm 1 Grass Dominant Color Extraction

Convert the input video file into its corresponding frames
for each frame **do**
    Convert the frame from RGB to HSI color space using equations (3), (4), and (5)

$$H = \cos^{-1}\left[\frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)+(G-B)]^{\frac{1}{2}}}\right] \tag{3}$$

$$S = 1 - \frac{3}{R+G+B} \times \min(R+G+B) \tag{4}$$

$$I = \frac{1}{3}(R+G+B) \tag{5}$$

    Define the color range that covers the different variations of the play-field's green color
    for Each pixel **do**
        if The three HSI components with range of values: $(0.15 < H < 0.4)$, $(0.1 < S < 1)$, $(0.2 < I < 1)$ **then**
            Set the color of this pixel to "$White$"
            Otherwise; set the color of this pixel to "$Black$"
        end if
    end for
end for

3.1.2 Shot boundary detection

Separated views come from multiple cameras positioned at different locations. It can be realized that while changing from one camera to another, this indicates a start and marks a boundary of a new shot. Accordingly, a shot can be defined as a sequence of frames recorded by a single camera with a continuous action in time and space [22]. There are two types of transitions depending on the camera movement and changing; namely, 1) *instant (cut) transition* and 2) *gradual transition*. Instant transitions are considered to be more accurate than gradual transitions [21]. Algorithm (2) shows the steps of shot boundary detection.

3.2 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.

3.2.1 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; *long*, *medium*, *close-up*, and *audience (out-of-field)* shot classes.

Algorithm 2 Shot Boundary Detection

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

An average dominant color ratio was used for all frames during the shot for determining the view-type of that shot. The shot-type classification algorithm proposed in this paper is based on a specific threshold (range) for grass ratio (G). It has been developed offline by training each class with different shots from different matches in order to define a specific range for each one of the four shot-types based on dominant color ratio histogram and assign each shot to one of the four different shot-types.

A set of thresholds have been defined for distinguishing the grass-ratio for the different shot-types [21]. For the proposed system, we applied four threshold ratios, each frame can be classified into one of the previously stated views depending on equation (6). Where, GL stands for grass ratio of long shot view, GM stands for grass ratio of medium shot view, GC/A stands for grass ratio of both close-up and audience (out of field) shot views. Furthermore, additional thresholds were used for distinguishing between close-up and audience shot views. BA was used as the average black color ratio in the whole frames during shot.

$$Shot.type = \begin{cases} Long.View, & G_L \geq 0.55; \\ Medium.View, & G_M \geq 0.15 \text{ and } G_M < 0.55; \\ Close.up.View, & G_{C/A} \leq 0.15 \text{ and } B_A < 0.8; \\ Audience.View, & G_{C/A} \leq 0.15 \text{ and } B_A \geq 0.8. \end{cases} \tag{6}$$

3.2.2 Play / break classification

For view-type classified shots, using start and end frames location of the shot, the boundaries of each play and break shot can be determined. The start of a play shot is identified as the first frame of long consecutive long-view frames, which can be interleaved by very short zoom-in or close-up shots. On the other hand, the start of a break shot is identified as the starting frame of either a long zoom-in shot or a shorter close-up shot, which can be interleaved by very short long-view shots. Consecutive play shots are considered as a play scene, which usually are ended with a consecutive break shots. Thus, a play-break

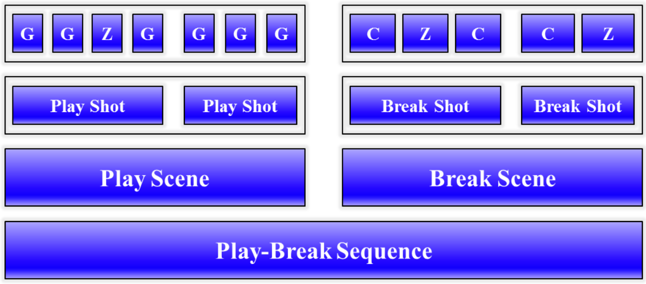

Figure 2. Play/break classification

sequence is a combination of consecutive play and break scenes, and sport games consist of many of this sequences [24]. The play-break actions are described in figure 2.

For shot-type classification, a set of thresholds have been defined for distinguishing the grass-ratio for the different shot-types [21]. For the proposed system, we applied four threshold ratios, each frame can be classified into one of the previously stated views [23]. On the other hand, for play/break classification, 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 [23, 24].

3.3 Replay Detection Phase

In most soccer matches, exciting events are often replayed. These exciting shots normally correspond to highlights in a game, e.g., actions near the goal posts in a soccer game. Replay is a video editing way that is often used to emphasize an important segment with a logo appearance for once or several times. The inputs to replay detection phase are the output shots from the shot boundary detection step of the pre-processing phase in order to extract the exciting events that are represented by the replay shots. 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 match. In the recent years, broadcasters use inserted logo sequence, as a digital video effect to replay the exciting and important events in soccer videos [25], as presented in figure 3 that shows an example of gradual logo appearance.

The aim of logo detection is the identification of a wide variety of logos used by different broadcasters. Based on experimental findings, a generic logo-model can be developed as: (1) *Logo* is meant to be contrasted from the background and is usually animated within 10-20 frames with a general pattern of *smallest-biggest-smallest* and (2) *Biggest contrast* usually takes at least 40-50 % of the whole frame, whereas smallest contrast is up to 10-20% [30].

For logo detection module, there are several processes passed by the input shot to determine the boundary of the replay scene, which is bounded by the logo. There is one feature, which is *Logo is appeared at the beginning of the shot*, generated by the shot detection stage

Figure 3. Gradual logo appearance

used in the logo detection module. That feature is used to enhance the logo detection performance as well as its accuracy because only few frames from the beginning of each shot will be processed. Algorithm (3) describes the steps of logo detection algorithm using support vector machine (SVM) and artificial neural network (ANN).

Algorithm 3 Logo Detection Using Different Machine Learning Classifiers

```
Train the ANN classifier with correct logo and false logo samples
Train the SVM classifier with correct logo and false logo samples
for Each frame do
  Adjust image intensity values for increasing the contrast of the input frame
  Select region of interest based on color for returning a binary image
  Calculate frame white ratio = the percentage of white pixels in the whole frame
  if Any frame contains a large contrast object (the white frame ratio be greater than
  0.5) then
    Get the original colored frame for the classification
    if The logo is real then
      Mark this shot as replay shot
    end if
  end if
end for
```

3.4 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 [5]. Excitement event detection is based on four features; namely, 1) *scoreboard detection*, 2) *vertical goal posts detection*, 3) *goal net detection*, and 4) *commentator loudness detection*.

3.4.1 Scoreboard detection

The scoreboard is a caption region distinguished from the surrounding region, which provides information about the score of the game or the status of the players [26]. 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 scoreboard is detected with enough confidence, it can undoubtedly provide the inference of goal event, because after every scored goal the scoreboard is displayed. The lower third of each frame was checked for containing a scoreboard via applying algorithm (3) as well.

3.4.2 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. Algorithm (4) presents the steps applied to each frame for detecting the vertical goal posts. Results of hough transform for detecting the vertical goal posts, as shown in figure 4.

Algorithm 4 Vertical Goal Posts Detection

```
for each frame do
    Use K-means clustering to convert each frame to binary image using squared Euclidean distances measure
    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 ≤ n) S = S_1, S_2, , S_k so as to minimize the within-cluster sum of squares (WCSS) using equation (7)
```

$$arg_S min \sum_{i=1}^{k} \sum_{x_j \in S_i} \parallel x_j - \mu_i \parallel^2 \quad (7)$$

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

Use Hough transform to detect the two goal posts using equation (8)

$$rho = x * cos(\theta) + y * sin(\theta) \quad (8)$$

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

```
    if The overlap between the vertical parallel lines greater than 80% then
        mark this frame as goal post frame
    end if
end for
```

3.4.3 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 [27]. The Gabor filter is used due to that the goal net has a unique

Figure 4. Hough transform for detecting the vertical goal posts

pattern and repeated many times.

The Gabor filter is basically a Gaussian filter, with variances sx and sy along x and y-axes, respectively. the sx and sy are modulated by a complex sinusoid, with center frequencies U and V along x and y-axes, respectively. The Gabor filer is described by using equations (9), (10), and (11).

$$G = \exp((\frac{-1}{2}(\frac{\acute{x}}{s\acute{x}})^2 + (\frac{\acute{y}}{s\acute{y}})^2) * \cos 2 * \Pi * f * \acute{x}) \tag{9}$$

$$\acute{x} = x * \cos(\theta) + y * \sin(\theta); \tag{10}$$

$$\acute{y} = y * \cos(\theta) - x * \sin(\theta); \tag{11}$$

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

3.4.4 Commentator 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 (12):

$$Volume = \frac{1}{N} * \sum_{n=1}^{N} |x(n)| \tag{12}$$

Where N is the number of frames in a clip and $x(n)$ is the sample value of the nth frame. To calculate pitch and silence, we applied the sub-harmonic-toharmonic ratio based pitch determination in [28] for its reliability. Louder, less silence, and higher pitch audio frames are identified by using dynamic thresholds presented in [29]. So, we can detect the excitement shots.

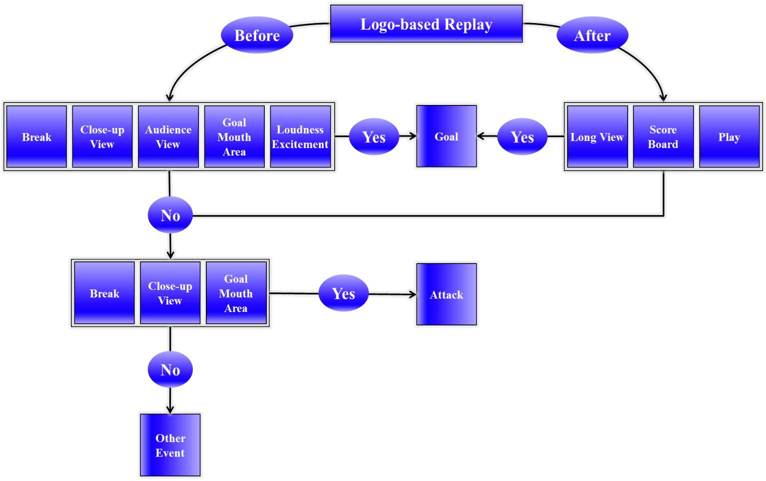

Figure 5. Event type classification

3.5 Event Detection and Summarization Phase

The summarized segment may contain only important events, such as: goal shots, attacks, or penalty shots [21]. 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 5 shows the different event type classification.

In this work, we evaluate a highlight with the following rules (sequence of features) for classify the events into goal, attack, and other events. Table 1 shows the cinematic features are used for each event detection.

3.5.1 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 [4]. Figure 6 illustrates the sequence of cinematic features after scoring a goal. Finally, the restart of the game is usually captured by a long shot.

As shown in table 1, the following features used to detect the goal event that occurs between the long shot resulting in the goal event and the long shot view that shows the restart of the game:

- **Replay Duration (RD)** due to goal events the duration no less than 20 and no more than 60 seconds.
- **Goal Mouth (GM)**
- **Close-up View (CV)**
- **Audience View (AV)**
- **Long View (LV)**

Feature	Description	Goal Event	Attack Event	Other Event
Replay Duration (RD)	Usually, the longer a replay scene is, the more attractive the segment is.	Yes	Yes	Yes
Goal Mouth (GM)	The goal mouth appearance are often shown before and within the replay scene in the cases of interesting goal, shoot and attack.	Yes	Yes	No
Close-up View (CV)	This is a shot of a close-up of a player who scored the goal.	Yes	Yes	Yes
Audience View (AV)	An excited audience shot will be displayed after an interesting event according to general video cinematic feature.	Yes	No	No
Long View (LV)	The long view to define the restart of the match after a replay (break).	Yes	Yes	Yes
Scoreboard Detection (SD)	This view are shown after a successful goal (a score). In these views, the caption containing score information is usually superimposed into long field-views.	Yes	No	No
Commentator Loudness (CL)	Loudness, silence and pitch generated by a commentator are effective measurements for detecting excitement.	Yes	No	No

Table 1. Event detection features

(a) long view of the actual goal play

(b) player close-up

(c) audience

(d) the first replay

(e) the second replay

(f) the third replay

Figure 6. An example of goal broadcast: the temporal order is from (a) to (f)

- **Scoreboard Detection (SD)**
- **Commentator Loudness(CL)**

3.5.2 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 [21]. 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. As shown in table 1, the following features used to detect attack event that occurs between two pair of replay logos:

- **Replay Duration (RD)** due to attack events the duration no less than 15 and no more than 35 seconds.
- **Goal Mouth (GM)**
- **Close-up View (CV)**
- **Long View (LV)**

Also, from table 1, the following features used to detect the other event that occurs between two pair of replay logos:

- **Replay Duration (RD)** due to other events the duration no less than no more than 15 seconds.
- **Close-up View (CV)**
- **Long View (LV)**

4 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 that is sampled at 44.1 kHz.

Results of shot boundary detection stage are shown in table 2. For a number of soccer match videos from the different championships with a total duration of 4hrs: 47min : 17sec, it has been obtained that 2081 shot boundaries have been correctly detected, while 194 shot boundaries have been failed to be correctly detected. These 194 shot boundaries have been divided into 135 falsely detected shot boundaries with addition to 59 missed shot boundaries. Accordingly, recall ratio for shot boundary detection was 97.2%, whereas precision ratio was 93.9%.

Table 2. Shot boundary detection results

Duration (hh:mm:ss)	Correct	False	Miss	Recall	Precision
4:47:17	2081	135	59	97.2 %	93.9 %

Table 3 and figure 7 shows the results of shot views classification stage. Recall ratios for long, medium, close-up and audience shot views classification have been obtained to be 91.8%, 87.1%, 85.7%, and 59.6%, respectively. Precision ratios for long, medium, close-up and audience shot views classification have been obtained to be 85.4%, 84.9%, 86.9%, and 85.3%, respectively.

Table 3. Shot classification results

Shot-Type	Long	Medium	Close-up	Audience
Long	925	97	46	15
Medium	83	807	17	19
Close-up	0	43	450	25
Audience	0	3	12	87
Recall	91.8%	87.1%	85.7%	59.6%
Precision	85.4%	84.9%	86.9%	85.3%

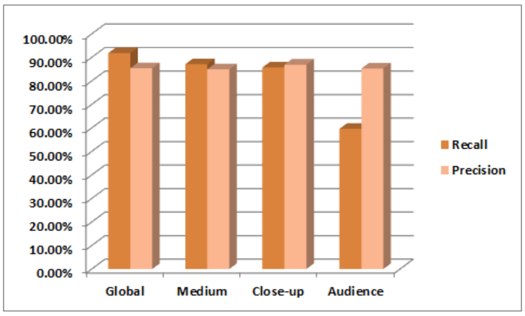

Figure 7. Shot classification results

Table 4 illustrates results of both SVM-based and ANN-based logo replay detection stage. Compared to the performance results obtained using SVM classifier, the proposed system attained good ANN-based performance results concerning recall ratio, however it attained poor ANN-based performance results concerning precision ratio.

Table 4. Evaluation of logo based replay using SVM and ANN

Factors	SVM	ANN
Duration (hh:mm:ss)	1:53:39	1:53:39
Correct	103	98
False	8	43
Miss	2	7
Recall	98.1%	93.3%
Precision	92.8%	69.5 %

Table 5 and table 6 show the results of scoreboard and goal mouth detection, respectively. For scoreboard detection, SVM classifier has been used whereas both Gabor filter and Hough transform have been used for goal mouth detection.

Table 5. Evaluation of scoreboard detection

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

Table 6. 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 7 and figure 8 shows the confusion matrix for event detection and summarization resulted from the proposed system.

Table 7. 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%

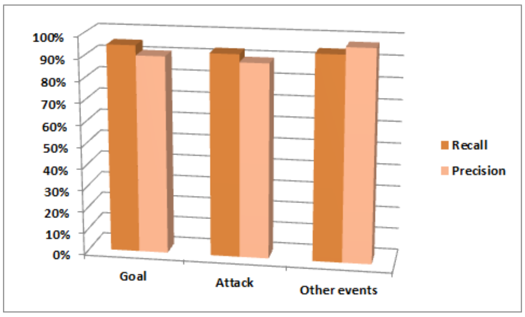

Figure 8. Results of event detection and summarization

5 Conclusions and Future Works

The ML-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 five phases; namely, pre-processing phase, shot processing phase, replay detection phase, excitement event detection phase, and 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 a high precision and reasonable recall ratios. Compared to the performance results obtained using SVM classifier, the proposed system attained good ANN-based performance results concerning recall ratio, however it attained poor ANN-based performance results concerning precision ratio. Accordingly, it has been concluded that using the SVM classifier is more appropriate for soccer videos summarization than ANN classifier. For future research, we will work on increasing the number of soccer videos and championships being examined in order to get more accurate results. Moreover, different machine learning techniques may be applied.

References

[1] Chen C.-Y., Wang J.-C., Wang J.-F., and Hu Y.-H., Motion Entropy Feature and Its Applications to Event-Based Segmentation of Sports Video. EURASIP Journal on Advances in Signal Processing, vol. 2008, Article ID 460913, 2008.

[2] D'Orazio T. and Leo. M., A Review of Vision-based Systems for Soccer Video Analysis. Pattern Recognition, vol. 43(8), pp. 2911-2926, 2010.

[3] Lotfi E. and Pourreza H.R., Event Detection and Automatic Summarization in Soccer Video. 4th Iranian Conference on Machine Vision and Image Processing (MVIP07), Mashhad, Iran, 2007.

[4] Ekin A., Tekalp A.M., and Mehrotra R., Automatic Soccer Video Analysis and Summarization. IEEE Transactions on Image processing, vol. 12(7), 2003.

[5] 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.

[6] Pan H., Li B., and Sezan M., Automatic Detection of Replay Segments in Broadcast Sports Programs by Detection of Logos in Scene Transitions. In: The IEEE International Conference on Acoustics, Speech, Signal Processing (ICASSP'02), Orlando, Florida, USA, pp. 3385-3388, 2002.

[7] Ancona N., Cicirelli G., Branca A., and Distante A., Goal Detection in Football by Using Support Vector Machines for Classification. In: The International Joint Conference on Neural Networks, Washington, DC, USA, vol. 1, pp. 15-19, 2001.

[8] Xu M., Duan L.Y., Xu C., Kankanhalli M., and Tian Q., Event Detection in Basketball Video Using Multiple Modalities. In: The Fourth International Pacific Rim Conference on Multimedia (PCM'2003), Singapore, vol. 3, pp. 15-18, 2003.

[9] Nepal S., Srinivasan U., and Reynolds G., Automatic Detection of Goal Segments in Basketball Videos. In: Proceedings of the ninth ACM international conference on Multimedia, Ottawa, Ontario, Canada, pp. 261-269, 2001.

[10] Ye Q., Huang Q., Gao W., and Jiang S., Exciting Event Detection in Broadcast Soccer Video with Mid-level Description and Incremental Learning. In: ACM International conference on Multimedia, Singapore, pp. 455-458, 2005.

[11] Rui Y., Gupta A., and Acero A., Automatically Extracting Highlights for TV Baseball Programs. In: Proceedings of the ACM international conference on Multimedia, Los Angeles, California, USA, pp. 105-115, 2000.

[12] Leonardi R. and Migliorati P., Semantic Indexing of Multimedia Documents. IEEE Multimedia, vol. 9, no.2, pp. 44-51, 2002.

[13] Assfalg J., Bertini M., Bimbo A.D., Nunziati W., and Pala P., Soccer Highlights Detection and Recognition Using HMMs. In: IEEE International Conference on Multimedia and Expo (ICME'2002), Lausanne, Switzerland, pp. 82582, 2002.

[14] Tovinkere V. and Qian R.J., Detecting Semantic Events in Soccer Games: Towards a Complete Solution. In: IEEE International Conference on Multi-media and Expo (ICME'2001), Tokyo, Japan, pp. 1040-1043, 2001.

[15] Tong X., Lu H., Liu Q., and Jin H., Replay Detection in Broadcasting Sports Video. In: Proceedings - Third International Conference on Image and Graphics, 2004.

[16] Kobla V. and Doermann D., Detection of Slow-Motion Replays for Identify Sports Videos. In: Proceedings of IEEE Third Workshop on Multimedia Signal Processing, pp. 135-140, 1999.

[17] Kobla V., DeMenthon D., and Doermann D., Identification of Sports Videos Using Replay, Text, and Camera Motion Features. In: Proceedings of the SPIE Conference on Storage and Retrieval for Media Database, vol. 3972, pp. 332-343, 2000.

[18] Yu B. and Zhu D.H., Automatic Thesaurus Construction for Spam Filtering Using Revised: Back Propagation Neural Network. Journal Expert Systems with Applications, vol.37(1), pp. 24-30, 2010.

[19] Bishop C.M., Neural Networks for Pattern Recognition. Oxford University Press, 1995.

[20] Wu Q. and Zhou D.-X., Analysis of Support Vector Machine Classification. J. Comput. Anal. Appl. vol. 8, pp. 99-119, 2006.

[21] Ekin A., Sports Video Processing for Description, Summarization and Search. PhD Thesis, University of Rochester, Rochester, 2003.

[22] Xing-hua S. and Jing-yu Y., Inference and Retrieval of Soccer Event. Journal of Communication and Computer, vol. 4(3), 2007.

[23] Zawbaa H.M., El-Bendary N., Hassanien A.E., and Yeo S.S., Logo Detection in Broadcast Soccer Videos Using Support Vector Machine. In: The 2011 Online Conference On Soft Computing in Industerial Applications WWW, (WSC16), 2011.

[24] 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.

[25] 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.

[26] 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(4), 2006.

[27] 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.

[28] 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.

[29] 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.

[30] Han B., Yan Y., Chen Z., Liu C., and Wu W., A General Framework for Automatic Online Replay Detection in Sports Video, In: Proceedings of the 17th ACM international conference on Multimedia (MM '09). ACM, New York, USA, pp. 501-504, 2009.