EVENT DETECTION IN SOCCER VIDEO USING BAYESIAN NETOWK

[1]Dharmesh R. Tank, [2]Dr Devang S. Pandya

[1] Assistant Professor, [2]Assistant Professor
[1]Information Technology,
[1]Silver Oak College of Engineering & Technology, Ahmedabad , India
[2]Information & Communication Technology,
[2]Adani Institute of Infrastructure Engineering, Ahmedabad , India

***Abstract*:-** Traditionally video were analyzed manually, however this costs valuable time. To sustain increasing growth of video data, there is an emerging demand. Therefore it is necessary to have an approach to make the video analysis task faster and automatically. Automatic video summarization is about extracting important events from original video to produce general summaries for the most important moments by interest. One way for creating a video summary is by extracting the features and then selecting key frames from these segment. Even though this approach is used for automatic video summarization, user cannot retrieve video intuitively, based on high level concepts. To solve this type of problem, event detection based video summarization is proposed in this paper. Proposed event detection approach aims to extract high level events like goal, player exchange and other types using Bayesian network.

***Keywords*:** Video Summarization, Key Frame, Video Skimming, Event Detection, Bayesian Network.

1. Introduction

Digital video provide better pictures than analogue. Digital video can be manipulated more easily than analogue video. In addition to this, digital video can be stored on random access media, whereas analogue video is generally stored sequentially on magnetic tape. This random access allows for interactivity, since individual video frames are addressable and can be accessed quickly. Digital video can be duplicated without loss of quality which is important for editing applications. However manipulating digital video is not exclusively preserve of film and television producers. Desktop video editing is possible on most high end desktop computers and many come with special hardware to digitize video. However, the user may not have sufficient time to watch the entire video (e.g. User may want to watch just the highlights of a game) or the whole of video content may not be of interest by the user (e.g. Replay video).

In such cases, the user may just want to view the summary of the video instead of watching the whole video. Moreover, in the advances digital cameras and camcorders have made it quite easy to capture a video and then load it into a computer in digital form. Many companies, universities and even ordinary families already have large repositories of videos both in analog and digital formats: such as the broadcast news, training and education videos, advertising and commercials, monitoring, surveying and home videos. All of these trends are indicating a promising future for the world of digital video.

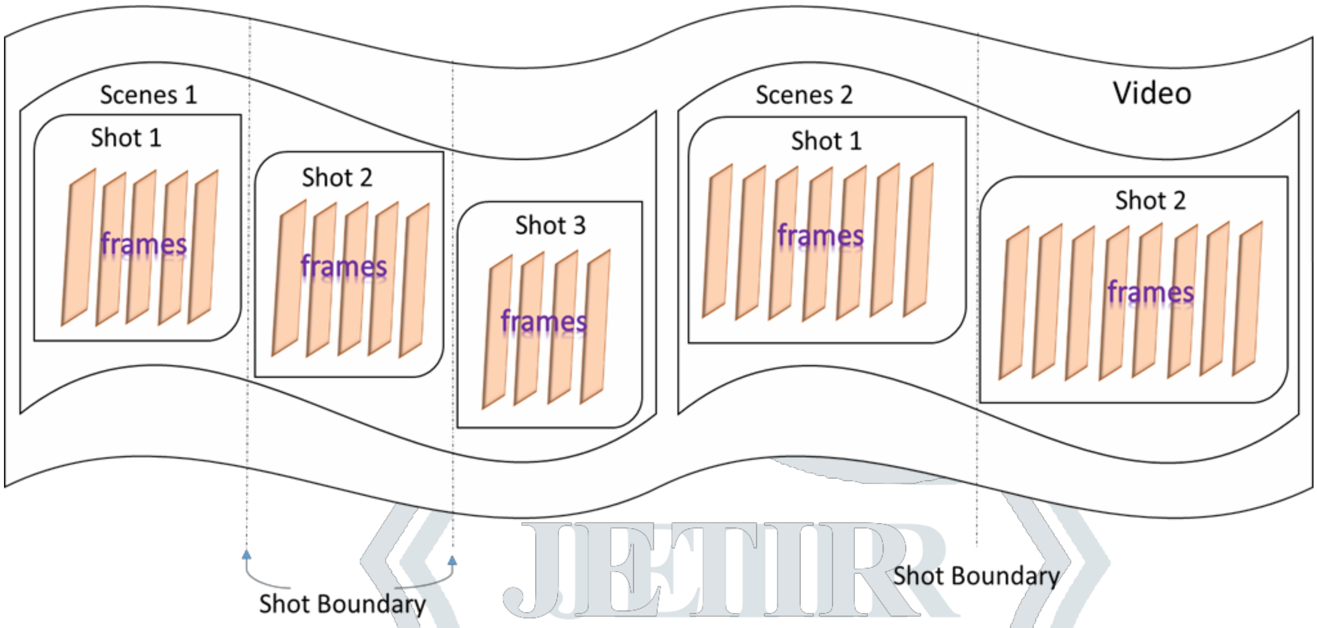

Figure 1: Frames Structure of Video

1.1 Video Summarization

Video Summarization is a short summary of the content of a longer video. It represents concise information about the content and also preserve essential message of original video data. It is a static visualization method, in which number of representative frames, often called key frames, are selected from the source video sequence. Video summary can be generated manually and automatically. Video summary is a collection of image sequences along with the corresponding audios from an original video sequence. Video skimming [1] is also called a preview of an original video, and can be classified into two sub-types: highlight and summary sequence. A highlight contains the most interesting and attractive parts of a video, while a summary sequence renders the impression of the content of an entire video. Among all types of video abstractions, summary sequence conveys the highest semantic meaning of the content of an original video. But as volume of the video data is high and manpower is limited fully automated solution is more preferable.

1.2 Video Events

The term "event" is used in many disciplines including mathematics, physics, and philosophy as well as in the terminology of culture and technology. "Events" as actions occurring in a linear time sequence. A video event understanding process takes an image sequence as input and abstracts it into meaningful units. The result of the abstraction is used by the event model to determine if an event of interest has occurred. Output of a video event understanding process may be a decision on whether a particular event has occurred or a summary of events in the input sequence. Automatic annotation systems are built so as to detect events of interest. Therefore we can firstly split events in interesting and non-interesting. In the case of video-surveillance interesting events can be specific events such as "people entering a prohibited area", "person fighting", "person damaging public property", or "person stealing something" and so on.

In the sport domain the detection of rare events is of interest, but systems need to detect events with a specific content (typically called highlights such as "scoring goal", "player injury", "yellow card", "player exchange", "goal attack" etc. Most of the domains in which video-analysis is performed involve the analysis of human motion (sports, video-surveillance, and movies). Features play a critical role in the analysis of video events. Good qualities are expected to be robust against variations in the same video of events in different positions would be recognized properly.

2. Literature Review

Various methods of video skimming and video summary are as discussed in sub sequent sections of this chapter. Most of video summarization techniques may be categorized into two classes: 1) Segmentation based 2) Event based.

2.1 Segmentation Based Video Summarization

It is defined as a set of key frames extracted from the video. This method can be applied to informative video content from all parts of the program can also be important for the user. Examples of the content of the video presentations, documentaries and trailer of films. This abstract may be a sequence of still images or moving images (video cropping). In general, content-based video summarization can be considered as a two stage process. The first step is to divide the video into frames, called segmentation videos or video shot detection limits. The second step is to find such a representative frame of each shot, which may well describe the video. Shot segmentation based crossover should get captured on video. Location change blew only require elaborate similarity metric between two frames. However, this simple process has some problems [8][9]. then another Cluster based approach it generated the cluster of such redundant frames and contain obtaining as the need of shot. A generalization of this method [10][11] depends to remove the visual content of premature termination of the video frames. All videos cluster nodes, each of which has the same framing pictures content. By representing each cluster and its representative body, a set of events frames in a given time, including the order.

2.2 Event Based Video Summarization

Event-based video summarization techniques are applicable to video contents that contain easily identifiable video units that form either a sequence of different events and non-events. For event based content, they are well defined by using knowledge based technique. If the application domain of the summarization algorithm is restricted to event-based content, it becomes possible to enhance summarization algorithms by exploiting domain-specific knowledge about events. Summarization of sports video has been the main application for such approaches. Sports programs lend themselves well for automatic summarization for following of reasons. Compact representations of sports programs have a large potential audience. Often there are clear markers, such as cheering crowds, stopped games, and replays, that signify important events. During any excitement event camera will move over different parts of the ground[5].

In most soccer videos, exciting events are often replayed to emphasize an important segment with a slow-motion pattern or logo appearance for one or several times. In broadcast sports video, replays provide the viewers another chance to watch the interesting events. The replays can be utilized for efficient navigation, indexing, and summarization of the sports video programs[12][13].

The score board provides information about the game and players thus provides an important cue for event detection as was used in. The caption often appears at the bottom part of image frame for a short while and then disappears. The score board detector relies on identifying the caption that appears at the bottom part of the frame and remaining present for a minimum duration[14].

Most of the exciting events are occur in the goal-mouth area which can be selected as highlighted candidates. For goal-post detection, first, we detect the goal post and its crossbar by searching for the white color, and then look for an intersection point between them. Appearance of goal mouth region in the frames will indicate attack or important situations.

The audience view is also very important with respect to detect the high semantic events. For audience view most of researcher are prefer the edges and texture of the frames. Audience shots always take place after a shooting or a verdict event. The attitude of the spectators infers the atmosphere of the stadium, and mark attracting events during a game[15][16].

The candidate events contain not only true exciting highlights but also some non-exciting highlights. However, it is not possible to identify the goal event. In the actual soccer games, goal events never appear in short sub-segments[17].

A useful way to consider player substitutions is as an intervention taking place at a specific point in time [18]. As a result of the intervention, the player moves to an alternative state, i.e., substitution. The concern therefore is the effect of the explanatory variable on the timing of the transition between states. Duration models allow us to determine the likelihood of a transition from one particular state to another, given the values of a series of explanatory variables.

3. Proposed Approach for Event Detection

Most of the techniques are domain specific and application dependent. In our proposed algorithm we aimed to classify the high impact events like as goal, player exchange and other events in soccer video. The proposed system uses frame based feature extraction approach rather than shot based approach.

3.1 Event Segmentation

Event segmentation is carried out by computing optical flow between consecutive frames of the video. As sports videos are very dynamic in nature and involve huge motion, optical flow becomes the most appropriate choice. We apply the Lukas and Kanade optical flow technique [19], which is a widely used differential method for optical flow estimation in computer vision. It is also less sensitive to image noise and also fast. The Lucas and Kanade method assumes that displacement of the image contents is approximately constant within a neighborhood (window) of the pixel under consideration [20].

Occurrence of any major event in soccer involves gathering of players, audience feelings and a rapid change in views. The camera undergoes huge motion during the occurrences of all major events in soccer. As the event occupies the time span in the video, it is necessary to distinguish the event boundary. However, this task is very challenging to identify such candidate frames which mark the beginning and ending of an event.

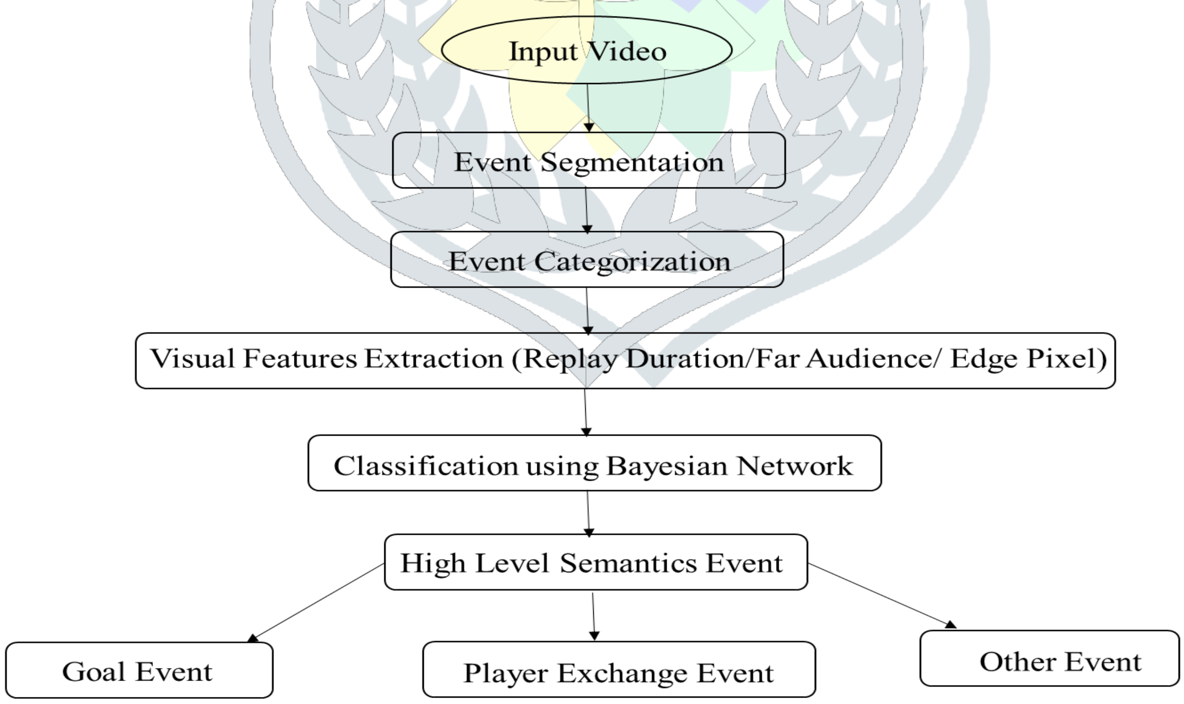

Figure 3.1: Block Diagram of Proposed System

3.2 Event Categorization

After segmentation of the events, we have to categorize these events. The event categorization phase splits the set of events into low-impact and high-impact sets. The high-impact set consists of events like goal, player exchange, injury etc. while the low-impact set includes events like goal attack, corner, foul, cheering in audience etc. Broadly, high impact events are longer in span while low-impact events are shorter.

3.3 Feature Extraction

In order to appropriately classify events, it is necessary to realize the temporal pattern of the frames of an event. To achieve this task, it is necessary to label every frame of the event of the video. This process is referred to as view classification. In order to carry out view classification, we extract visual features from the frames and classify them into different views. In the proposed method we can read multiple frames of a specific video. Convert all the frames from RGB to HSV frames. HSV values are so important to find out the green pixels. After finding the green pixels values, we need to find the dominant grass pixel ratio by using those green pixel ratio. Clustering can be done by K-Means clustering method for the view classification.

- **Far- Audience View:**
  Far and audience views are the semantically most important features to be used for event detection. There are many view classification are illustrate in [12] by the researcher. Among them as per our system requirement we are going to compile the Far view and Audience view. Most goal events are followed by a celebration which involves gathering of players and cheering in the audience, so one can easily look out for this feature. In our system, we are introduce a new feature by combining these two. Goal event involve most number of different views. In that goal event involve far as well as audience views. When the goal attack events resultant into a goal event, the celebration of the player as well as audience drag the attention due this characteristics. At the end of goal event, camera focuses on the audience views to show the cheering in the audience. This is the reason that we utilize the far view and audience view in order to analyse the goal event. Player exchange event is semantically quite different than the goal event. This event narrate the replacement of the player and hence concern keeps focus on the particular player who is entering or either leaving the ground. Due to this characteristic, most of the views fall in the category of close or far views. Hence the dominance of the far audience views is lesser in player exchange compared to the goal event.

- **Replay Duration:**
  Replay of any events provide a clear idea that some interesting events took place during the match. It depend upon the importance of the events that how long replay should be. Generally in any sports video, broadcaster or provider generates its own logo for replay occurrence. Logo frames will be shown at the beginning of the replay and ending of it span. In our case goal event and player exchange events produced more excitement and fast variation of the special features. In Goal event, the replay duration contain more frames span (>500 frames) than any other events and for player exchange event, replay duration will not been considered because for player exchange event does not require any replay. Hence we used this feature to easily classify between goal and player exchange event. For replay duration we proposed an algorithm as shown below:

Algorithm 1: Replay Duration through Template Matching

1: Select logo template which contains clear logo in the centre of the frame
2: Convert RGB resized frames into HSV images.
3: Convert RGB resized frames into YCbCr images.
4: Extract the Hue and Y Components from each frame

$$Average = \sqrt{H^2 + Y^2} \quad (1)$$

5: Repeat until event ends

i. Read all the frames and resize
ii. Repeat step 2 and 3 for the resized frames
iii. If difference between Hue and Y components average value with the earlier Hue and Y components are near to zero than template matched and replay duration ratio will be calculate as :-

$$RDR = \frac{Number\ of\ frame\ between\ two\ logo\ frames}{Maximum\ Span\ of\ Replay\ Duration} \quad (2)$$

6: Stop

Replay is depend upon different type of logo object provide by different broadcasters. In our approach we find out the replay duration ratio for classify the high impact events. Above algorithm provide an efficient way to achieve the replay duration ratio. We converted some logo frames as template into HSV as well as YCbCr color models. Through that averaging of Hue and Y components values are used for comparison with other logo frames. If the template matched with the experiment logo frames than the average value will near to null. After computing this we move towards our feature that is replay duration ratio.

According to the equation 4, number of frames between two logo is divide by the maximum number of span of replay frames, provide the replay duration ratio. This feature is directly classify the high impact events in our model as goal event and player exchange event. Because player exchange event does not required any replay after the event take place. It is observed that no broadcasters are showing the replay of any player exchange event during match. On the other hand goal event consist comparatively largest replay duration with different view side of cameras.

- **Edge Pixel Ratio:** Edge pixel provide a crystal view of any objects to identify. After the occurrence of goal events and player exchange, every broadcaster displays the caption about the goal information and player substitute information, respectively. Broadcasters display the caption containing the information of players who has scored the goal and his team name just after the occurrence of an event. This caption almost stays for 4 to 8 second. After the completion of an event, the broadcaster displays the caption of team score information mostly at the bottom part of the frame. For Player exchange the broadcasters display a caption for in player as well as out player. It is seen that green color represent the entry of the player in match and red color denote out from the match. Apart from this, in player exchange event, team member is also showing a number board to display the substitute the player during the match.

Generally, it is observed that the goal score caption is displayed within 55 second after the goal event and for player exchange it will remain for 15-20 second. Hence for player exchange this feature is play an important role. The presence of such views and detailed caption becomes a very important clue for the confirmation of the goal event as well as for player exchange. We do not consider initial frames up to 4 second (100 frames) of an event for EPR computation, and we continue to compute EPR for another 55 second (1,250 frames) even after the end of an event for the inclusion of the goal score caption. We follow some algorithmic steps to carry out edge analysis of high-impact events.

Algorithm 2: Edge Caption and Pixel Ratio

1. Initially taking an image and resized it.
2. Divide the image horizontally into three equal parts and choose the bottom part where the caption is located.

3. Convert the image in gray and apply the Canny Operator to detect horizontal edges.
4. Compute EPR

$$EPR = \frac{Total\ Number\ of\ edge\ pixels}{Total\ number\ of\ pixels\ in\ the\ frames} \tag{3}$$

5. Stop

We convert the original image into gray scale image. Generally we are observed that most of broadcaster provide their caption at below part of image. We experiment on different types of captions in different ground fields' condition. Then we moved to edge detection from the frames by applying canny edge detection. This operation provide all the edges within the image. By applying morphological operation (Erode) horizontal lines are detected. The dense of the edges provide more ideas about the caption identification. For caption we use the feature canny edge detection.

The caption is also displayed for a longer duration while the player is leaving and a new player is entering. When the player leaves the ground, a red triangular symbol is displayed within the caption, and a green triangular symbol is displayed within the caption while a new player is entering the ground. This caption contains the important triangular shape of either red or green color. The EPR of every frame of an event is computed, and then we calculate the average value of the EPR of an event and also the average EPR of all high-impact events. Canny operator is an edge detection operator that uses a multi-stage algorithm to detect a wide range of edges in images [22][23]. By partitioning the bottom of the frame for edge detection. We extract only horizontal lines for caption which are generally below within a frame.

3.4 Event Classification using Bayesian Network

In order to classify the events from a video, state event models can be used; such as SVM, Bayesian classifier, Neural network, Decision tree and HMM. Among these, the most powerful tool that can be used as a classifier for semantic video analysis is the Bayesian Inference. BN is powerful semantic analysis tools which have been applied to model the high-level semantic information embedded in the video data. Here, we use BN to model the semantic highlights of soccer game.

- **Bayesian Network:** A Bayesian network model representation of a problem domain can be used as the basis for performing inference and analyses about the domain. Decision options and utilities associated with these options can be incorporated explicitly into the model, in which case the model becomes an influence diagram, capable of computing expected utilities of all decision options given the information known at the time of decision. Bayesian networks and influence diagrams are applicable for a very large range of domain areas with inherent uncertainty.

The Bayesian network is a graphical model for Bayesian classifier which is called an optimal classifier because of its minimum classification error.

$$P(X1, X2, \ldots, Xn) = \prod_{i=1}^{n} P(Xi \mid Parents(Xi)) \tag{4}$$

$$P(C|X) = P(X1|C) * P(X2|C) * \ldots \ldots \ldots * P(Xn|C) * P(C) \tag{5}$$

Where

$P(C|X)$ is the posterior probability of class (target) given predictor (attribute).

$P(C)$ is the prior probability of class.

$P(X|C)$ is the likelihood which is the probability of predictor given class.

$P(X)$ is the prior probability of predictor.

- **BN Learning and Inference:** Fitting the graphical models is called learning. In many practical settings the BN is unknown and one needs to learn it from the data. The first step consists in finding the BN graph structure that encodes the conditional independencies present in the data. It should at least identify a distribution as close as possible to the correct one in the probability space. This step is called network structure or simply structure learning. The second step is called parameter learning, as the name suggests, deals with the estimation of the parameters of the global distribution.

  The BN learning problem [29], which can be stated informally as follows: Given training data and prior information (e.g., expert knowledge, casual relationships), estimate the graph topology (network structure) and the parameters of the joint probability distribution(JPD) in the BN. Learning the BN structure is considered a harder problem than learning the BN parameters. Moreover, another obstacle arises in situations of partial observability when nodes are hidden or when data is missing. In general, four BN learning cases are often considered, to which different learning methods are proposed, as seen in Table 3.1. In the first and simplest case, BN is known and the learning method is maximum likelihood estimation, find the values of the BN parameters with Conditional Probability Density (CPD) that maximize the (log) likelihood of the training dataset. The BN is automatically generated after the following training process rather than determined by ad hoc.

- **BN with Discrete and Continuous Variables:** BN for large and complex domains are difficult to construct, maintain and also time consuming. Therefore to manage such type of complex structure network, we prefer two approaches as discrete and continuous variables [30]. If all the random variables in X are discrete (Bottom up approach) or categorical then both the global and the local distributions are assumed to be multinomial. Bayesian networks is best model for it, because of its strong ties with the analysis of contingency tables and it allows an easy representation of local distributions as conditional probability tables.

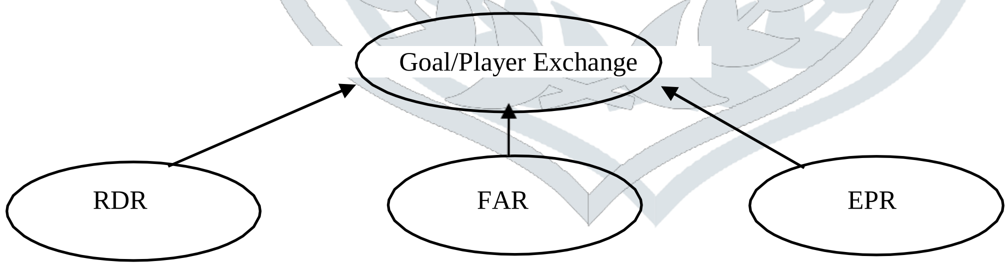

Figure 3.9 Bottom up Approach for Discrete Variables

If all the random variables in X are continuous, the global distribution is usually assumed to follow a Gaussian distribution, and the local distributions are either univariate or multivariate Gaussian distributions. This assumption defines a subclass of graphical models called graphical Gaussian models (GGMs).

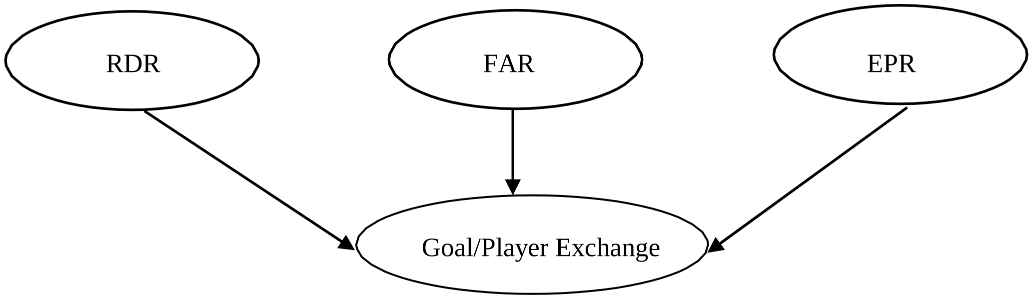

Figure 3.10 Top down Approach for Continuous Variables

If both continuous and categorical variables are present in the data there are three possible choices: assuming a mixture or conditional Gaussian distribution.

Inference procedures for Bayesian models focus mainly on evidence propagation and model validation, even though other aspects such as robustness and sensitivity analysis have been studied for specific settings. Evidence propagation studies the impact of new evidence and beliefs on the parameters of the model. For this reason it is also referred to as belief propagation or belief updating and has a clear Bayesian interpretation in terms of posterior and conditional probabilities. The structure of the network is usually considered fixed, thus allowing a scalable and efficient updating of the model through its decomposition into local distributions. New evidence is introduced by either altering the relevant parameters (soft evidence) or setting one or more variables to a fixed value (hard evidence). The former can be thought of as a model revision or parameter tuning process, while the latter is carried out by conditioning the behaviour of the network on the values of some nodes.

4. Simulation Results and Discussions

In the experiments, football video were recorded from TV broadcasting programs. During the simulation we used 11 datasets of varying length. The resolution of the video frame is 352 * 288 and the frame rate is 25 frames per second. We have experimented with soccer videos of total length almost 6 h. We have conducted experiments on 11 video datasets from seven well-known soccer leagues like Barclays Premier League, La Liga, Serie A Premier League, FIFA Euro Cup, Europa, England 2, NPower League and Champions League. All these videos possess varying ground and illumination conditions, e.g. daylight, floodlight as well as shadow. Video dataset information is shown in Table 4.1 along with the video illumination condition. Date information of the match is shown in DD/MM/YYYY format. We can easily observe in each of the video the far and close view is very different from other videos. This is normal because of different leagues videos and a variety of illumination conditions. We aim to propose a novel Bayesian model for a large number of various leagues and every type of features. Experimental results are evaluated using standard parameters: precision and recall. Precision quantifies what proportion of the detected events is correct while recall quantifies what proportion of the correct events is detected. If we denote D the events correctly detected by the algorithm, D_m the number of missed detections (the events that should have been detected but were not) and D_f the number of false detections (the events that should not have been detected but were), we have:

$$\text{Recall} = D / (D+D_m) \tag{6}$$

$$\text{Precision} = D / (D+D_f) \tag{7}$$

Video No.	Match Information (Team Name, League Name, Date, Condition)	Duration Minutes
1	Germany vs Greece, 2nd half, FIFA Euro cup 2012, 21/6/2012, floodlight	38:00
2	Cardiff vs Middlesbrough, 1st half, England 2, 2/5/2011, daylight	35:00
3	FC Barcelona vs Getafe, 1st half, La Liga 2011, 19/3/2011, floodlight	10:00
4	Porto-Spartak Moscow, 2nd half, Europa League,40640, floodlight	40:00
5	Real Sociadad vs Athletic Club, 2nd half, La Liga, 29/9/2012, floodlight	33:00
6	Real Madrid vs Deportivo, 1st half, La Liga,41182, floodlight	22:00
7	Catania vs Intermilan, Serie A, 1st half, 15/10/2011, floodlight	16:00
8	Arsenal vs Montpellier, 2nd half, Champions League, 21/11/2012, floodlight	46:00
9	Southampton vs Spurs, 1st half, Barclays Premier League, 22/12/2013, daylight	48:00
10	Celta Vigo vs RCD Mallorca, 2nd half, La Liga,41231, daylight	20:00
11	Hundersfield vs Peterborough, 1st half, NPower League one playoff, 29/5/2011, daylight	48:00

Table 4.1 Soccer Video Dataset

4.1 Classification Results using Discrete Variables

Here are the results of our experiments over discrete variables. The classification results of discrete variables are shown in table 4.2.

Video No	Total Goals	Correct (D)	False (D_f)	Missed (D_m)	Precision (%)	Recall (%)
1	5	1	1	4	50	20
2	2	0	1	2	0	0
3	1	1	0	0	100	100
4	4	2	2	2	50	50
5	1	1	1	0	50	100
6	2	2	0	0	100	100
7	1	1	0	0	100	100
8	1	0	2	1	0	0
9	1	1	3	0	25	100
10	1	0	1	1	0	0
11	0	0	3	0	0	0
Total	19	9	14	10	39.13043	47.36842

Table 4.2 Goal Event Classification Result for Discrete Variables

Video No	Total Goals	Correct (D)	False (D_f)	Missed (D_m)	Precision (%)	Recall (%)
1	2	1	3	1	25	50
2	0	0	2	0	0	0
3	0	0	0	0	100	100
4	0	0	0	0	100	100
5	3	2	0	1	100	66.66667
6	0	0	0	0	100	100
7	0	0	0	0	100	100
8	3	2	0	1	100	66.66667
9	0	0	1	0	0	0
10	0	0	2	0	0	0
11	0	0	0	0	100	100
Total	**8**	**5**	**8**	**3**	**38.46154**	**62.5**

Table 4.3 Player Exchange Event Classification Result for Discrete Variables

Video No	Total Goals	Correct (D)	False (D_f)	Missed (D_m)	Precision (%)	Recall (%)
1	3	2	2	1	50	66.66667
2	4	1	2	3	33.33333	25
3	1	1	0	0	100	100
4	4	2	2	2	50	50
5	2	2	0	0	100	100
6	0	0	0	0	100	100
7	0	0	0	0	100	100
8	3	2	1	1	66.66667	66.66667
9	5	1	0	4	100	20
10	5	3	0	2	100	60
11	10	7	0	3	100	70
Total	**37**	**21**	**7**	**16**	**75**	**56.75676**

Table 4.4 Other (Thrown in, Goal attack, Injury, Offside) Events Classification Result for Discrete Variables

4.2 Classification Results using Continuous Variables

According to our classification results, for continuous variables provide much better results than discrete variables. There is a strong reason behind it, because discrete module directly classify the events into binary way. Either event take place or not. But in continuous module classification can generated in most precise way. Because the presence of the mid-level features provides more probability to occur an event optimally.

Video No	Total Goals	Correct (D)	False (D_f)	Missed (D_m)	Precision (%)	Recall (%)
1	5	4	0	1	100	80
2	2	1	0	1	100	50
3	1	1	0	0	100	100
4	4	4	0	0	100	100
5	1	1	0	0	100	100
6	2	2	0	0	100	100
7	1	1	0	0	100	100
8	1	1	0	0	100	100
9	1	1	1	0	50	100
10	1	1	3	0	25	100
11	0	0	0	0	100	100
Total	**19**	**17**	**4**	**2**	**80.95**	**89.47**

Table 4.5 Goal Event Classification Result for Continuous Variables

Video No	Total Goals	Correct (D)	False (D_f)	Missed (D_m)	Precision (%)	Recall (%)
1	2	2	0	0	100	100
2	0	0	1	0	0	0
3	0	0	1	0	0	0
4	0	0	1	0	0	0
5	3	3	0	0	100	100
6	0	0	0	0	100	100
7	0	0	0	0	100	100
8	3	3	3	0	50	100
9	0	0	0	0	100	100

10	0	0	1	0	0	0
11	0	0	1	0	0	0
Total	**8**	**8**	**8**	**0**	**50**	**100**

Table 4.6 Player Exchange (PE) Event Classification Result for Continuous Variables

Video No	Total Goals	Correct (D)	False (D_f)	Missed (D_m)	Precision (%)	Recall (%)
1	3	3	1	0	75	100
2	4	3	1	1	75	75
3	1	0	0	1	0	0
4	4	3	0	1	100	75
5	2	2	0	0	100	100
6	0	0	0	0	100	100
7	0	0	0	0	100	100
8	3	0	0	3	0	0
9	5	4	0	1	100	80
10	5	1	0	4	100	20
11	10	9	0	1	100	90
Total	**37**	**25**	**2**	**12**	**92.59**	**67.56**

Table 4.7 Other (Thrown in, Goal attack, Injury, Offside) Events Classification Result for Continuous Variables

As shown in the both comparisons graph, we observed that continuous module contain high preciseness and high recall in different analytical conditions. Discrete module consist the results in binary mode to classify any events. Hence it generates less precision and recall compare to its opposite module.

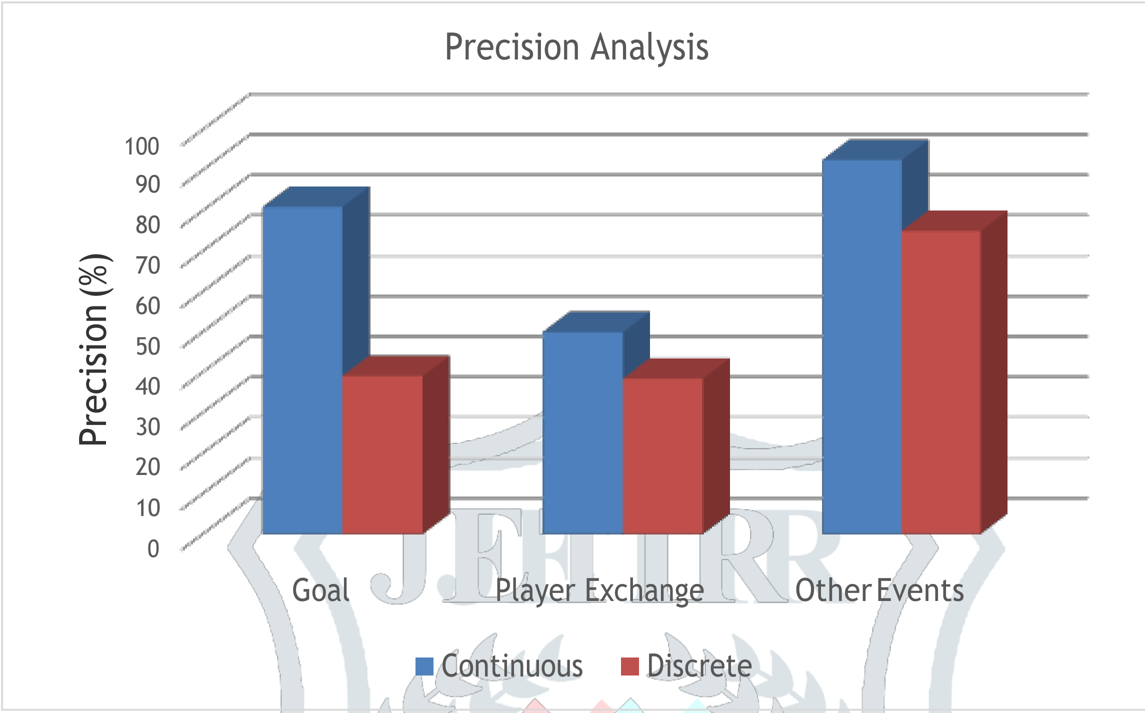

Figure 4.1 (a) Comparative Performance for Precision

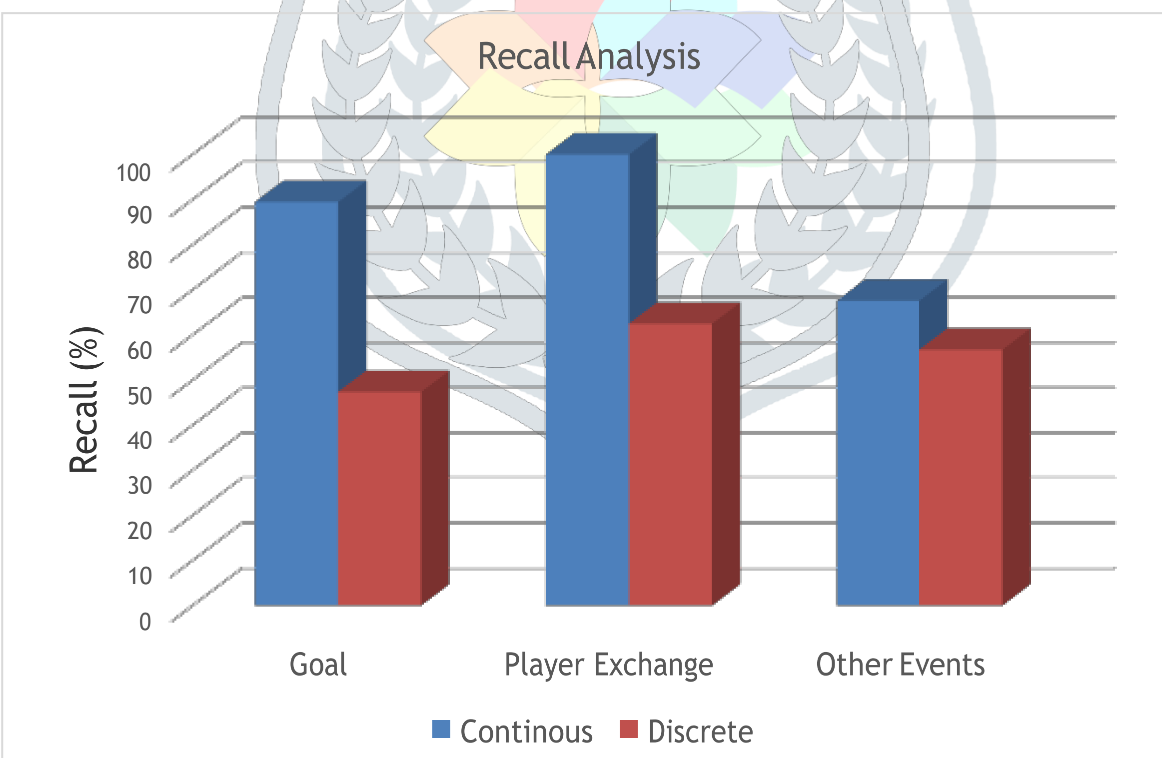

Figure 4.1 (b) Comparative Performance for Recall

5. Conclusion

Event based video summary is very essential in the navigation or retrieval of large amount of video collections available on the internet. Video summarization attempts to answer the problem of dealing with large amount of video by creating abstract of these videos. Sport videos specifically soccer video are very exciting and full of enthusiasm hence attract major viewers across the world. People love to watch some interesting segments of matches repeatedly because of their interest rather than watching the entire match. Analysis of the match is very important for the coaches as well as player. Hence event based video summary is the essential for the players and coaches to understand the strengths and weaknesses. Most of the analysis carried out on sport videos are domain specific and dependent on the particular game.

We aim to propose the algorithm for the detection of high energetic events like goal, player exchange and other types of events using Bayesian network. We use mid-level features like replay duration, far-audience ration, and edge pixel ratio to understand the energetic events. These features are formulated using low level features like color, and edges. Bayesian structure is a powerful mathematical tool which efficiently combine the graph and probability theory. Bayesian structure is learnt using the derived mid-level features.

We have tested our algorithm on various types of datasets which include variety of leagues from different broadcasters. Algorithm classifies the goal event and player exchange with high precision and recall. Looking at the various leagues, it is clear that algorithm is robust as it successful operates on matches having varying conditions. Algorithm successful also classifies the other events with high precision and recall. We have experimented with mid-level features modelled as continuous and discrete. It is found that classification accuracy with continuous features is more than the discrete features. This is obvious because events are random in nature. Hence features of an event also behave randomly. Our experiments clearly reflect that modelling mid-level features as Gaussian continuous random variables is advantageous and favourable.

References

[1] M. H. Kolekar, K. Palaniappan, S. Sengupta, and G. Seetharaman, "Semantic Concept Mining Based On Hierarchical Event Detection for Soccer Video Indexing," *J. Multimed.*, vol. 4, no. 5, pp. 298–312, 2009.

[2] A. Ekin, M. Tekalp, and R. Mehtrotra, "Automatic Soccer Video Analysis and Summarization," *IEEE Trans. Image Process.*, vol. 12, no. 7, p. 12, 2001.

[3] H. M. Zawbaa, N. El-bendary, A. E. Hassanien, and T. Kim, "Event Detection Based Approach for Soccer Video Summarization Using Machine learning," *Int. J. Multimed. Ubiquitous Eng.*, vol. 7, no. 2, pp. 63–80, 2012.

[4] D. Tjondronegoro, Y.-P. P. Chen, and B. Pham, "Sports Video Summarization using Highlights and Play-Breaks," in *MIR '03: Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieval*, 2003, pp. 201–208.

[5] G. Lavee, E. Rivlin, and M. Rudzsky, "Understanding Video Events: A Survey of Methods for Automatic Interpretation of Semantic Occurrences in Video," 2009.

[6] Y.-G. Jiang, S. Bhattacharya, S.-F. Chang, and M. Shah, "High-level Event Recognition in Unconstrained Videos," *Int. J. Multimed. Inf. Retr.*, vol. 2, no. 2, pp. 73–101, 2012.

[7] S. F. De Sousa Junior, A. A. De, and D. Menotti, "An Overview of Automatic Event Detection in Soccer Matches," in *2011 IEEE Workshop on Applications of Computer Vision, WACV 2011*, 2011, pp. 31–38.

[8] C. Cotsaces, N. Nikolaidis, and I. Pitas, "Video Shot Detection and Condensed Representation - A Review," *IEEE Signal Process. Mag.*, vol. 23, no. 2, 2006.

[9] G. S. Zhao, "A Novel Approach for Shot Boundary Detection and Key Frames Extraction," in *Proceedings - 2008 International Conference on MultiMedia and Information Technology, MMIT 2008*, 2008, pp. 221–224.

[10] A. Sony, K. Ajith, K. Thomas, T. Thomas, and P. L. Deepa, "Video Summarization by Clustering using Euclidean Distance," *2011 Int. Conf. Signal Process. Commun. Comput. Netw. Technol.*, pp. 642–646, 2011.

[11] T. Chheng, "Video Summarization Using Clustering," *Dep. Comput. Sci. Univ. Calif.*, 2007.

[12] H. Pan, B. Li, and M. I. Sezan, "Automatic Detection of Replay Segments in Broadcast Sports Programs by Detection of Logos in Scene Transitions," *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4. pp. IV–3385–IV–3388, 2002.

[13] D. L. G. Krvvdp, D. Prwd, and J. Frp, "Soccer Video Summarization Using Enhanced Logo Detection Mohamed Y . Eldib , Bassam S . Abou Zaid , Hossam M . Zawbaa , Mohamed El-Zahar , Motaz El-Saban Information Technology department , Faculty of Computers and Information , Cairo University ," *Event (London)*, pp. 4345–4348, 2009.

[14] S. U. N. Xing-hua and Y. Jing-yu, "Inference and Retrieval of Soccer Event," vol. 4, no. 3, 2007.

[15] C. Huang, H. Shih, and C. Chao, "Semantic Analysis of Soccer Video using Dynamic Bayesian Network," *IEEE Trans. Multimed.*, vol. 8, no. 4, pp. 749–760, 2006.

[16] F. Tian and P. Shi, "Attractive Events Detection in Soccer Videos Based on Identification of Shots," *Int. Conf. Multimed. Technol.*, pp. 814–822, 2013.

[17] H. Kim, S. Roeber, A. Samour, and T. Sikora, "Detection of Goal Event in Soccer Videos," *Methods*, vol. 5682, pp. 317–325, 2005.

[18] J. Del Corral, C. P. Barros, and J. Prieto-Rodriguez, "The Determinants of Soccer Player Substitutions: A Survival Analysis of the Spanish Soccer League," *J. Sports Econom.*, vol. 9, no. 2, pp. 160–172, 2007.

[19] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Proc. 7th Int. Jt. Conf. Artif. Intell.*, pp. 674–679, 1981.

[20] D. S. Pandya and M. a Zaveri, "A Novel Framework for Semantic Analysis of an Illumination-Variant Soccer Video," *EURASIP J. Image Video Process.*, vol. 2014, no. 1, p. 49, 2014.

[21] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An Efficient K-Means Clustering Algorithm: Analysis and Implementation," vol. 24, no. 7, pp. 881–892, 2002.

[22] S. Saini, B. Kasliwal, and S. Bhatia, "Comparative Study Of Image Edge Detection Algorithms," 2013.

[23] R. Maini and H. Aggarwal, "Study and Comparison of Various Image Edge Detection Techniques," *Int. J. Image Process.*, vol. 147002, no. 3, pp. 1–12, 2009.

[24] R. C. Gonzalez, R. E. Woods, and B. R. Masters, *Digital Image Processing, Third Edition*, vol. 14, no. 2. Prentice Hall, 2009.

[25] Z.-N. Li and M. S. Drew, *Fundamentals of Multimedia*, 2nd ed. 2004.

[26] U. B. Kjærulff and A. L. Madsen, “Probabilistic Networks - An Introduction to Bayesian Networks and Influence Diagrams,” no. May, 2005.

[27] K. P. Murphy, “An Introduction to Graphical Models,” pp. 1–19, 2001.

[28] S. Alipour and P. Oskouie, “Bayesian Belief Based Tactic Analysis of Attack (Events in Broadcast Soccer Video),” *Int. Conf. Informatics, Electron. Vis.*, pp. 612–617, 2012.

[29] B. Networks, F. Faltin, and R. Kenett, “Bayesian Networks,” *Encycl. Stat. Qual. Reliab.*, vol. 1, no. 1, p. 4, 2007.

[30] M. Scutari and K. Strimmer, “Introduction to Graphical Modelling,” in *Handbook of Statistical Systems Biology*, 2011, pp. 235–254.

[31] C. Fox, “Bayes Net Toolbox Practical.” p. 5, 1995.