EURASIP Journal on Image and Video Processing
a SpringerOpen Journal

RESEARCH Open Access

A novel framework for semantic analysis of an illumination-variant soccer video

Devang S Pandya[1*] and Mukesh A Zaveri[2]

Abstract

This paper presents an effective, novel and robust framework for semantic analysis of a soccer video possessing varying illumination conditions. The proposed algorithm works in two phases. The proposed framework effectively detects and gathers important events in the first phase, and a later phase carries out the task of event classification. The proposed system aims to identify high-level semantics of a soccer video like card event, goal event, goal attack and other classes of events. The proposed framework effectively exploits optical flow and colour features to detect and classify the events. The use of event filtration and categorization features successfully makes the system effective over various conditions. Simulations have been performed on a large number of video datasets of different conditions of various soccer leagues. Simulation results reflect the efficiency and robustness of the proposed framework.

Keywords: High-level semantics; Optical flow; Event filtration; Event categorization

1 Introduction

Rapid revolution in digital video has brought many applications at home in affordable cost. The volume of digital data has been increasing rapidly due to the wide usage of multimedia applications in the areas of education, entertainment, business, medicine etc. One of the major applications is sports video analysis. Sports videos attract majority of people due to their capability of producing thrill as well as uncertainty of results. Hence, there has been an enormous increase of such video contents on the Internet. Video summarization helps to address these needs by developing a condensed version of full-length videos [1-3]. Extraction of important events and creation of summaries do not only make the video compact but also make it possible to deliver over low-bandwidth networks.

A raw video is an unstructured data stream, physically consisting of a sequence of video shots. A video shot is composed of a number of frames, and its visual content can be represented by key frames. Most of the video summarization techniques may be categorized into two classes, segmentation-based video summarization and event-based video summarization. The former is defined as a collection of key frames extracted from a video. This type of technique is applicable to uniformly informative video content where all parts of the programme may be equally important for the user. Examples of such contents are presentation videos, documentaries and home movies. This summary can be a sequence of stationary images or moving images (video skims). In general, content-based video summarization can be thought as a two-step process. The first step is partitioning the video into shots, called video segmentation or video shot boundary detection. The second step is to find such representative frames from every shot that can well describe the video. Thus, the video can be considered as a collection of shots and every shot consists of key frames. 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. Such kind of summary is usually presented as an organized sequence of interesting events. The best example of this type of video content is sports highlights. In sports highlights, exciting events such as wicket fall, hitting a six in cricket, goal, issuing a card in soccer etc. are included and dull segments are eliminated. Other examples are surveillance videos, talk shows and news programmes. The past several years have observed significant research to the event-based

* Correspondence: devang.pandya@ganpatuniversity.ac.in
[1]U V Patel College of Engineering, Ganpat University, Kherva 384012, Mehsana, India
Full list of author information is available at the end of the article

video summarization of various kinds of sports such as soccer, baseball, tennis, cricket etc.

Event-based video summarization extracts the highest meaningful contents of the video which is the most favourable in the end-user perspective. In [4], different solutions to video summarization have been described in detail. Low-level features such as colour, motion and textures are important and widely used features for such video processing. There exist different colour histogram-based approaches in which consecutive frames are compared to decide key frames. In [5,6], HSV colour space is used to measure interframe difference. It has been shown that the HSV (hue, saturation, value) colour space has outperformed the RGB (red, green, blue) colour space due to its perceptual uniformity. RGB mutual information and joint entropy of adjacent two frames have been used in [7] for marking key frames. A colour histogram is insensitive to camera and object motion. Therefore, colour-based key frame selection may not be able to deliver the visual content of the shot. Motion is an important clue about camera zooming and panning. To address camera motion, optical flow components are extracted and the motion function is computed between two frames in [8-10].

Object motion trajectories and interactions have been used for soccer play classification and for soccer event detection [11-13]. However, both [11] and [12] rely on pre-extracted precise object trajectories, which were generated manually in [11] and are not useful for real-time applications. For a soccer video, rule-based classification has been used in [12]. Xie et al. have carried out classification by defining mutually exclusive states of the game, namely, play and break [14]. A combination of cinematic and object descriptors has been used in [15]. A superimposed caption embedded in the video has been used in [16] to detect the scoreboard of the baseball videos. Semantic features along with replays and audio energy have been applied for soccer video summarization in [17]. In [18], rule-based approaches have been used for the detection of events in sports videos.

Several approaches use stochastic methods that employ self-learning capability to extract knowledge such as hidden Markov models (HMM). HMM has been extensively used to detect events of the different types of sports videos [19-22]. A knowledge-based semantic inference has been applied for event recognition in sports videos in [23]. A combination of speech-band energy and pattern recognition based on dominant colour provides important clues for event detection in soccer videos [24]. Researchers have also attempted dynamic Bayesian network (DBN) which is based on the Bayesian network (BN) and its extension [25]. Learning through DBN has been applied to many real applications [26,27]. In [27], the Bayesian belief network has been used for the analysis of goal attack events of soccer videos. In the last recent years, machine learning-based approaches have dragged attention. A machine learning system that learns to predict video transitions based on feature information derived from the frames is developed. In supervised learning, low-level features are employed to train the system that can predict transitions on unseen data [28,29]. Various unsupervised learning approaches have been attempted in the literature to extract key frames and found faster as they do not require training [30]. An adaptive neuro-fuzzy system has been proposed in [31]. Literature study reveals that stochastic and machine learning approaches have drawn the major attention of researchers. Sports videos experience huge motion so it is a natural cue, and we exploit the motion as a low-level feature for our proposed framework. The rest of the paper is organized as follows. The proposed framework is described in Section 2. Section 3 discusses experimental results and Section 4 concludes the paper.

2 Proposed framework

Semantic video analysis involves the inclusion or identification of the major events of the video. We propose a robust and efficient framework for semantic analysis of a soccer video possessing highly varying illumination conditions. It is robust in the sense that it succeeds to achieve favourable results even under various conditions of the soccer video with minimal assumptions. It is also efficient as we apply low-level features like colour and motion to detect the important events while neglecting object-based features which are computationally expensive. The proposed framework of semantic analysis of a soccer video is shown in Figure 1. The proposed framework is fully automatic. The entire video is processed frame by frame. The block diagram of the framework is briefly described below.

1) The first step of our framework is to carry out event segmentation. We propose a novel algorithm for event segmentation based on change in optical flow between the successive frames of the video. Change in the horizontal component of optical flow is found very successful to segment the video. At the end of event segmentation, we have a set of events. There are many dull events which are not important to the end users. A block diagram trifurcates after the event segmentation. Each process is carried out independently than the others which is subsequently described.
2) We propose a novel and robust card event detection algorithm. The algorithm understands the caption and exploits this domain knowledge effectively to detect the card event. After event segmentation, the card detection algorithm is applied on the obtained set of events.

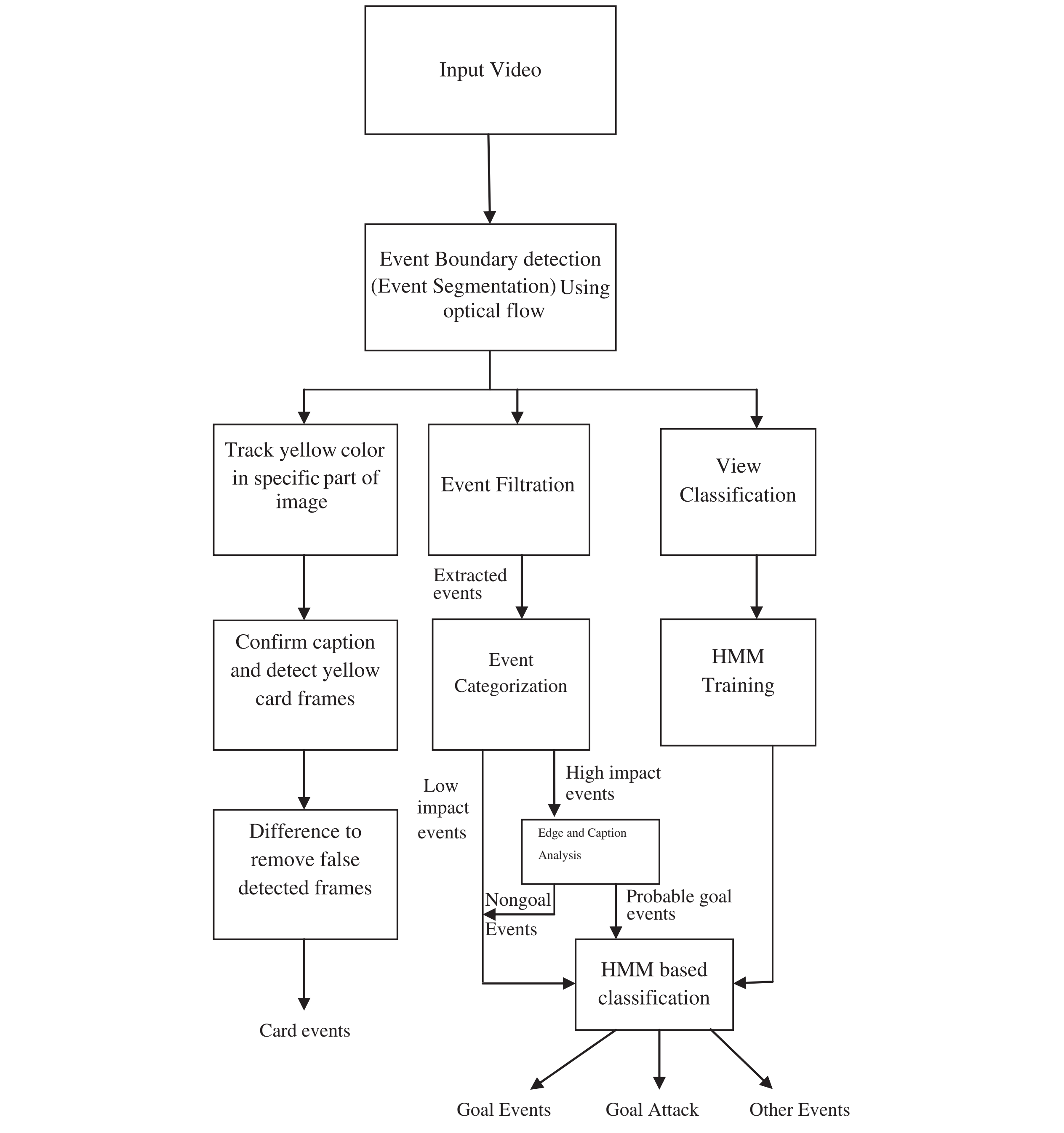


Figure 1 The proposed framework of semantic analysis of a soccer video.

3) Event filtration is applied to remove dull events like just passing the ball on the ground, audience views etc. After obtaining meaningful events, event categorization is applied to separate out high-impact and low-impact classes of events. High-impact class of events is lengthy as well as involve more view transitions. High-impact class consists of events like goal, injury, player exchange etc. while low-impact class consists of events like goal attacks or corner and other events like throw in, offside etc.
4) View classification is carried out for all the detected event segments. This algorithm heavily depends on the dominant colour and edges. The algorithm is robust to the varying conditions of the ground which extremely affect the grass colour. Grass colour exhibits different brightness in flood light condition compared to the daylight condition. Based on this

view classification, HMM models are generated for the goal, corner/goal attack and other types of events.

5) Event classification is carried out on the low-impact and probable goal class of events using the obtained HMM models.

All the above-mentioned steps of the framework are described in detail in the subsequent Section 2 of the paper.

2.1 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 [9], 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 neighbourhood (window) of the pixel under consideration. The velocity vector (V_x, V_y) must satisfy:

$$Av = b \tag{1}$$

where:

$$A = \begin{bmatrix} I_x(q_1) & I_y(q_1) \\ I_x(q_2) & I_y(q_2) \\ \cdot & \cdot \\ I_x(q_n) & I_y(q_n) \end{bmatrix}, \quad b = \begin{bmatrix} \text{-}I_t(q_1) \\ \text{-}I_t(q_2) \\ \cdot \\ \text{-}I_t(q_n) \end{bmatrix}, \quad v = \begin{bmatrix} V_x \\ V_y \end{bmatrix}$$

q_1, q_2,..., q_n are pixels inside the window, and $I_x(q_i)$, $I_y(q_i)$ and $I_t(q_i)$ are the partial derivatives of image I with respect to positions x, y and t evaluated at pixel q_i and at the current time. The solution of Equation 1 is obtained by the least squares method. It is computed as:

$$v = \begin{bmatrix} \sum_i I_x(q_i)^2 & \sum_i I_x(q_i)I_y(q_i) \\ \sum_i I_x(q_i)I_y(q_i) & \sum_i I_y(q_i)^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i I_x(q_i)I_t(q_i) \\ -\sum_i I_x(q_i)I_t(q_i) \end{bmatrix}$$

In the experiments, the neighbourhood (window) size is set to 3. For the group of N +1 frame, N optical fields (F_1, F_2,..., F_n) will be computed by the algorithm. Before applying the optical flow computation algorithm, the resolution of the frames is down-sampled by a factor of 2 to speed up the computation. After obtaining the optical flow component, the optical flow magnitude is computed using Equation 2. It is observed that optical flow components V_x and V_y are quite sensitive to the shot transition also. This is natural due to global camera motion.

$$M(i) = \sum\nolimits_{(\mathrm{x,y})\in \mathrm{F_i}} \sqrt{{V_x}^2(x,y) + {V_y}^2(x,y)} \tag{2}$$

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. Camera motion is well and effectively observed by the optical flow components. We have emphasized on change in the horizontal component V_y as cameras track the soccer ball which has a more horizontal movement. We propose a novel feature to measure optical flow variation. For this, we differentiate Equation 2 with respect to V_y. The obtained equation is described below:

$$M_y^{'}(i) = \frac{\sum\nolimits_{(x,y)\in F_i} \sqrt{V_y^2(x,y)}}{\sum\nolimits_{(x,y)\in F_i} \sqrt{V_x^2(x,y) + V_y^2(x,y)}} \tag{3}$$

The above equation is found very efficient to exhibit a noticeable change at the beginning or at the end of an event period. As the event occupies the time span in the video, it is necessary to demarcate the event boundary. However, this task is very challenging to identify such candidate frames which mark the beginning and ending of an event. Figure 2 clearly reflects the occurrences of major events in soccer by showing a larger fluctuation in $M_y^{'}$. We can easily understand that in the case of a goal event, $M_y^{'}$ undergoes a rapid and large change due to frequent shot transition and rapid camera motion. Threshold is decided automatically to detect important events by using the following min-max normalization equation:

$$T = \frac{\bar{M}' - \mathrm{min}_{M'}}{\mathrm{max}_{M'} - \mathrm{min}_{M'}} (\mathrm{new\_max}_{M'} - \mathrm{new\_min}_{M'}) + \mathrm{new\_min}_{M'} \tag{4}$$

where $\bar{M}'$ is the average value of $M_y^{'}$, $\mathrm{min}_{M'}$ is the minimum value of $M_y^{'}$, $\mathrm{max}_{M'}$ is the maximum value of $M_y^{'}$. new_max and new_min values have been set to 0.5 and 0.8, respectively. After performing a large number of experiments on various video datasets, it has been observed that the minimum value of the M' is 0.5 for the detection of any event. Based on a newly determined threshold (T), significant events are demarcated. Important events like card, corner and goal continue over a certain minimum time span.

Using this fact, we consider only those events which sustain more than 5 s. This fact helps to reduce false detections. In Figure 2, the horizontal line is the threshold which is calculated using Equation 3. The computed $M_y^{'}$ value is varying over time (frames). As shown in Figure 2, it consists of multiple peaks corresponding to a significant soccer event. Figure 3a,b shows the corner and goal event sequences, respectively. These sequences are the series of various views like goal post views, close-up,

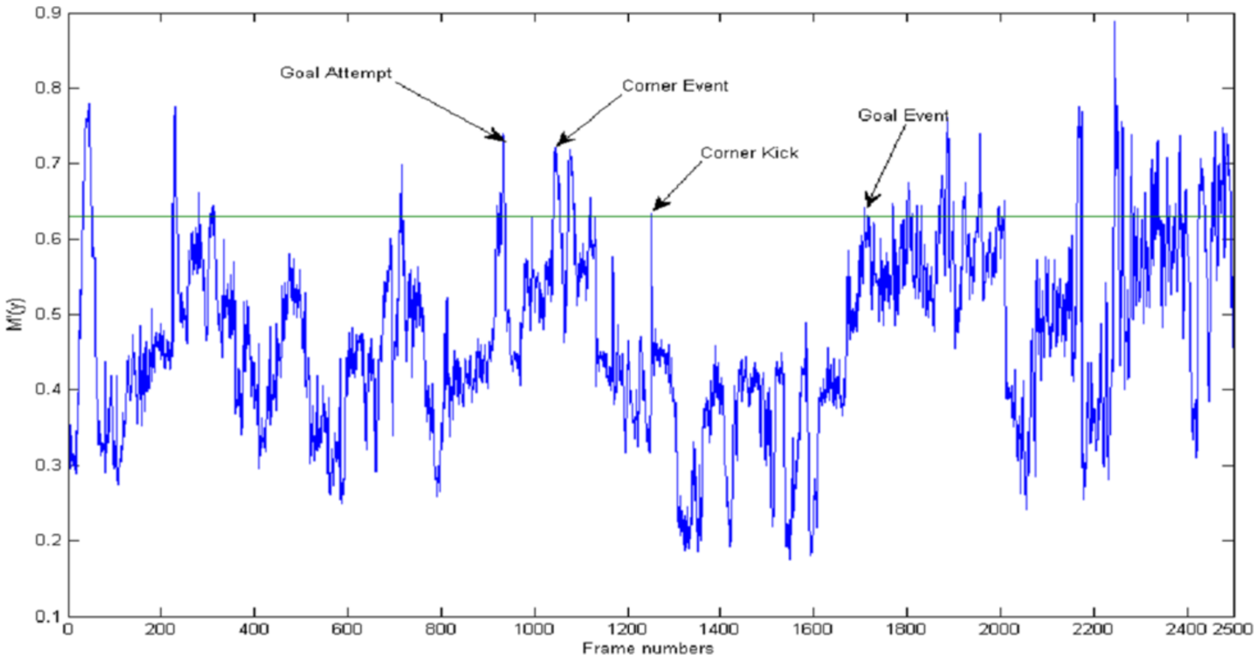


Figure 2 Change in horizontal optical flow component according to Equation 3.

player gathering etc. As the goal event shows a large number of continuous higher peaks above the threshold, it clearly exhibits more fluctuations compared to the corner event and it also continues longer than the corner event. Any $M_y^{'}$ which is higher than the obtained threshold marks the candidate frames which indicate the presence of an event. To separate out two events, there is at least a gap of 6 s between two peaks of $M_y^{'}$; otherwise, the next peak also contributes to the first event. This is valid because the corner event lasts at least for 5 s while the card and goal events are much longer than the corner event.

2.2 Yellow card event detection

Soccer is an eventful sport. Any unfair behaviour or some foul may cause the issue of a yellow card to the player. The yellow card itself is a very important event because it

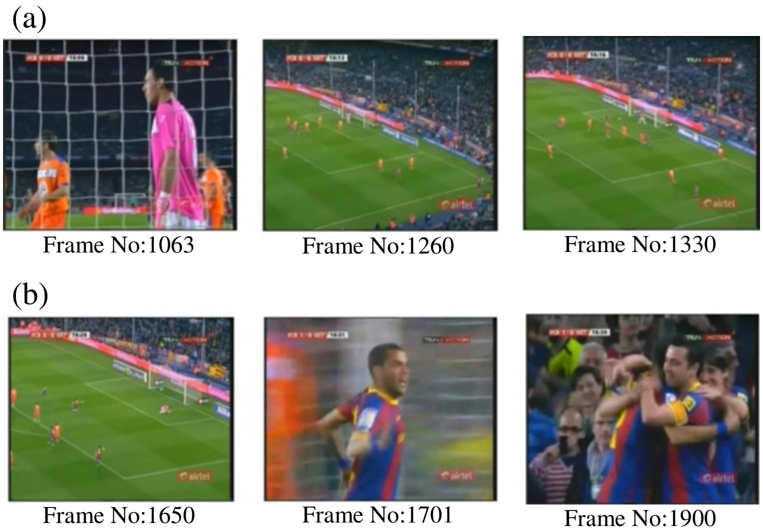


Figure 3 Event sequences: depicting transition of views. (a) Corner event sequence. **(b)** Goal event sequence.

is like a warning to the issued player. Issuing a second yellow card to the player (red card) compels the player to discontinue the game. This event itself gains very high importance; if the yellow card has been issued in the penalty area, then the opponent may be offered a penalty kick which may eventually end in scoring a goal by the opponent. Hence, it may generate a series of events. An algorithm to detect a yellow card by examining every frame of the segments obtained after event segmentation is described below.

Normally, a yellow card caption stays 2 to 4 s on the screen while the whole event lasts longer. The caption remains on display for the duration of almost 50 to 100 frames. A novel algorithm to detect a yellow card frame is proposed, which is described below. Generally, the yellow card is displayed as a caption with the player information at the bottom part of the image. It is observed that in most broadcast soccer videos, the caption appears on the bottom of the screen. We process an input frame below the centre of an image and that is also not exactly at the bottom area as it is very difficult to have an idea of the caption placement and its size and height variations. This domain knowledge helps us to detect this event easily rather than classifying this event from a set of events like goal, corner and even other types of events.

Yellow card frame detection algorithm

Phase I

```
1. Divide the image horizontally into three equal parts
   and choose the bottom part where the caption is
   located.
2. Convert the input frame from the RGB image into a
   grey image.
3. Apply a canny edge detection operator on the grey
   image to detect horizontal lines of the rectangular box.
4. Erode the horizontal lines which have a length
   shorter than the threshold.
5. Extract the connected components.
6. If number components >1, then
   a. Convert the image from the RGB image into an
      HSV image.
   b. Set the pixels with the highest grey level values
      (255) which have the hue, saturation and value
      range as per empirically defined values. Set the
      rest of the pixels to grey value 0.
   c. Convert the image in binary.
   d. Extract the connected components.
      i. If any connected component is found which
         has the number of pixels in a specified range,
         1. Declare a probable yellow card frame.
      ii. Otherwise, neglect and proceed to the next
          frame.
   e. Otherwise, neglect and proceed to the next
      frame.
7. Collect probable yellow card frames.
8. Keep those frames as yellow card frames if the
   minimum number of frames within that segment
   meets the defined criteria.
```

Phase II

```
9. Compute the edge pixel ratio of the cropped image.
   Apply the Sobel operator to find edges.
10. If the edge pixel ratio (EPR) is within the threshold,
    declare that segment as a yellow card event.
```

Phase III

```
11. Perform steps 1 to 6d on two chosen yellow card
    frames of every segment.
12. Find the absolute distance between the detected
    connected component.
13. If the distance is less than the threshold.
    a. Declare a yellow frame and accept the segment.
    b. Else, discard the segment.
```

The entire algorithm involves three phases. The first phase consists of one to eight steps which find probable yellow card frames based on the presence of yellow colour. Steps 1 to 3 are straightforward. Step 4 involves erosion which uses a horizontal structuring element of size 20 which removes all lines shorter than a length of 20. In step 6, we first check for connected components because their absence indicates a smooth or constant-intensity image. The display of a yellow card in the caption exhibits various colours, hence gives rise to a number of edges. After steps 3 and 4, we still have few longer edges left which eventually contribute to connected components. Step 6a does the conversion of the RGB image to HSV as an HSV model is considered perceptually uniform. We deal with yellow cards of largely varying shades, so HSV becomes the most appropriate model. The pure yellow colour is represented at 60° in HSV; this defines a range of hue for yellow colour at 0.16 (60°/360°). But there may be variation in the intensity of yellow colour in different league videos. It is observed that the hue range for yellow colour at 0.14 to 0.22 is found to be satisfactory for detecting yellow card frames. From an empirical study, the threshold values of the saturation and value components of HSV are set greater than or equal to 0.8 and 0.6, respectively. These thresholds are set in step 6b of the algorithm. At the end of step 6b, we confirm the presence of yellow colour and we set the yellow region with the highest grey level intensity while the rest of the pixels are set with the lowest intensity. In step 6c, the binary threshold has been set to 0.8 because a strict threshold removes all the unnecessary components. From observation of different broadcast videos of standard league matches, the area of the yellow card is fixed between 20 to 450 number of pixels (step 6d) which represent a smaller to wider size of yellow card at a different tilt.

As the yellow card stays for 2 to 4 s, step 8 uses this knowledge to identify the yellow card frames. The minimum number of frames which is required to declare a yellow card event is set to 15. Various types of yellow cards are shown in Figure 4. Figure 4 clearly depicts the largely varying size, intensity as well as location of the yellow card. Figure 5 shows various intermediate steps of the yellow card event detection algorithm. Few leagues display a yellow card whose intensity keeps varying. The second phase attempts to mark the presence of caption in these detected probable yellow card frames. In soccer videos, there are misleading frames like a player wearing a yellow t-shirt or yellow socks and even a yellow ball. We compute the edge pixel ratio of the cropped image to confirm the presence of a caption showing a yellow card. The edge pixel ratio is computed using following equation:

$$\mathrm{EPR} = \frac{\text{Total number of edge pixels}}{\text{Total number of pixels in the frame}} \qquad (5)$$

The range of EPR is set between 0.38 and 0.55. However, every event is narrated with the support of the caption, so there could be misleading cases where the frame shows a player wearing a yellow t-shirt along with the caption which conveys the information of a goal event. Phase III takes care of misleading frames that have a caption which coincides with a yellow t-shirt or some yellow logos of the t-shirt. This type of typical case is shown in Figure 6. In the case of wrong yellow event frames, due to the movements of the player, the player's t-shirt of these frames experiences motion. Step 11 selects two frames at some specific interval of the detected yellow event segment. If these are genuine yellow event frames, then after applying steps 1 to 6d, almost similar images are produced consisting of one connected component at the place of the yellow card. We find the city block distance between the top left coordinates of the connected components of both images. Absolute distance is the sum of x and y differences of the top left coordinate of connected components. Due to motion, the city block distance between both images will be high.

2.3 Event filtration

After event segmentation of the video, we obtain the set of events. The change in optical flow is a key parameter for demarcation of events in an input video. As we are processing a broadcast video, there is no control over capturing of video content, i.e. the camera is moved over the ground from one angle to another and it results into a change in optical flow and leads to an event segmentation which does not actually represent any event. We apply event filtration on this set to filter out certain events which are not significant and dull according to the interest of the end user. In order to carry out this task, we apply Fourier transform on $M_{y}^{'}$. Every event is characterized by the mean of the magnitude of the Fourier transform. Fourier transform and the magnitude can be found using the following formula. If the event is short and not important and there are smooth and

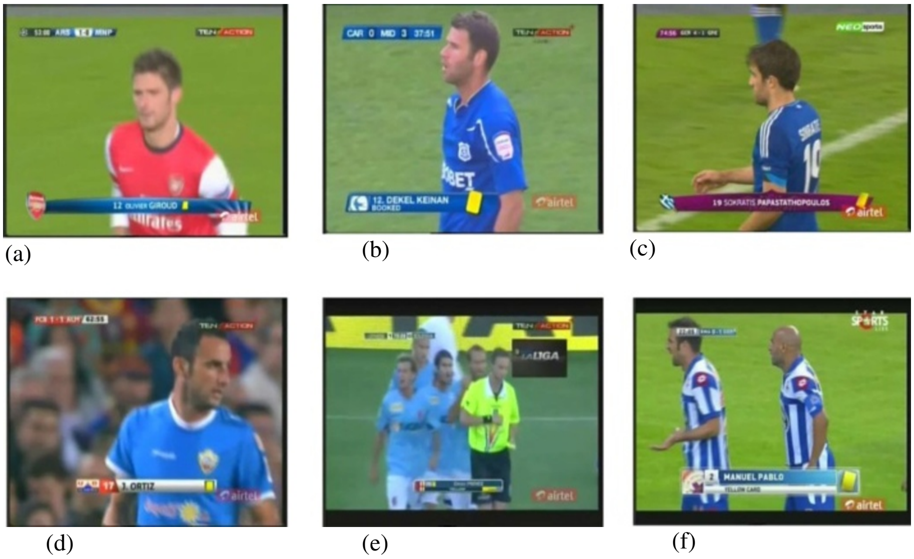


Figure 4 Various types of yellow cards. (a) Video 13, frame no. 11053. **(b)** Video 9, frame no. 48230. **(c)** Video 11, frame no. 35620. **(d)** Video 3, frame no. 18680. **(e)** Video 7, frame no. 5968. **(f)** Video 5, frame no. 2580.

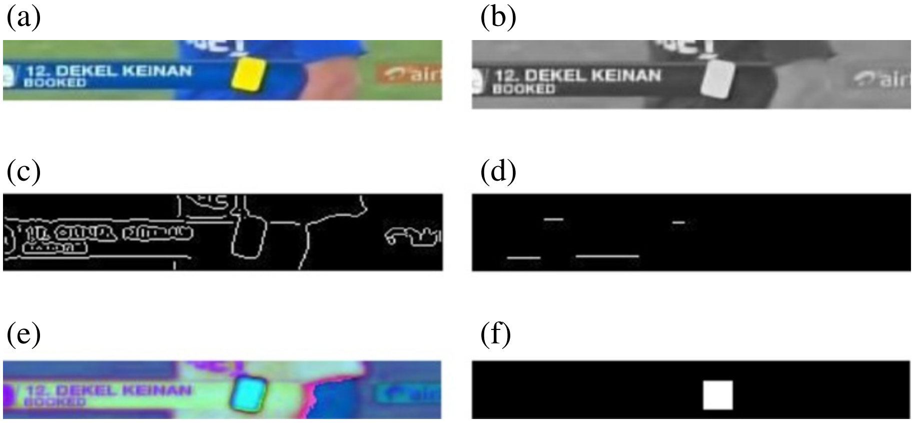


Figure 5 Intermediate steps of the yellow card event detection algorithm. (a) Cropped image. **(b)** Gray image of **(a)**. **(c)** Canny edge detection. **(d)** Erosion of image **(c)**. **(e)** HSV image of **(a)**. **(f)** Binary image of **(e)** after step 6c.

fewer fluctuations within the event, it has a smaller magnitude of Fourier transform. Even if the event is short in time duration but consists of many views like far, goal post, audience etc. and transitions among these views, then it gives rise to the magnitude of high-frequency components. This can be very well captured using Fourier transform magnitude. We choose only 10% of frequency coefficients and extremely high-frequency components.

$$\Im\left(M_y^{'}(u)\right) = \frac{1}{N}\sum_{x=0}^{N-1} M'(x)e^{-j2\pi ux/N}$$

$$|\Im(M')| = \sqrt[2]{R^2(u) + I^2(u)} \tag{6}$$

However, shorter events can also be an important event, so every event can be further analysed by how many frames of an event experience the change in average motion magnitude ($M_y^{'}$) above the threshold. This is a useful descriptor because goal and player exchange events last longer and also involve frames which have a greater change in motion than the threshold. This can easily be found by the formula given in Equation 7.

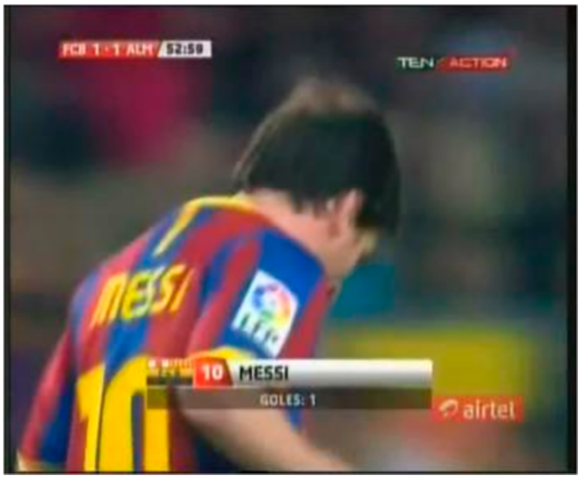


Figure 6 Caption along with yellow colour in the background.

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. However, there are goal events which are shorter and may not be followed by the much more cheering and celebration, but still due to more camera movements, they involve more frames undergoing a larger change in motion. We successfully think to use the product of the two features: Fourier transform and the ratio of frames having a change in motion greater than the threshold to total frames of an event. This product feature itself carries the neutral effect of Fourier transform and change in motion greater than the threshold. We introduce this product as an event filtration feature (EFF). This product feature enhances the capability to filter out insignificant events. We compute the mean of the EFF of every event. Events which succeed to satisfy the following criteria will be selected as filtered events:

$$\Pr\left(M_y^{'} > T\right) = \frac{\text{Number of frames whose } M_y^{'} > T}{\text{Total number of frames in an event}} \tag{7}$$

$$\text{EFF}_i > \alpha_1 \times \overline{\text{EFF}} \tag{8}$$

where $\overline{\text{EFF}}$ is the average value of the EFF of every event while α_1 is the empirical parameter which can be set between 0 and 1. We have set the value of α_1 to 0.7. EFF_i corresponds to the EFF value of event i. Next, we proceed to the event categorization phase.

2.3.1 Event categorization

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. Goal event is the most valuable event for the end users as well as for the game itself. For each event, the following features will be computed. Each event is characterized by the n number of $M_y^{'}$ values. Using these values, the first kurtosis is computed for every event. It is a descriptor of the shape of a probability distribution. Higher kurtosis means more of the variance is the result of infrequent extreme deviations. Goal event produces more deviations which can be frequent or infrequent. We can conclude that for the goal event, the value of kurtosis cannot be low but the values will be more than average or high. Kurtosis is computed using the following equation:

$$K_i = \frac{E\left(M_y^{'} - \mu\right)^2}{\sigma^4} \tag{9}$$

where μ is the mean and σ is the standard deviation of the $M_y^{'}$ values. Second, we compute the energy for each event i using the following sum of squared formula:

$$e_i = \sum M' y^2 \tag{10}$$

The above equation has quite good capacity to realize the event which is longer over time span and having higher $M_y^{'}$ values. In soccer, goal and player exchange types of event can be easily identified by this parameter. To address this issue, we formulated an event categorization feature (ECF), which is a product of kurtosis and the energy of an event:

$$\mathrm{ECF}_i = K_i \times e_i \tag{11}$$

High-impact events are selected using following equation:

$$\begin{aligned}\text{High\_impact} = &\left(F_{\text{EVENT}} > \alpha_2 \times \overline{\text{ECF}}\right) \\ &\text{and } \left(F_{\text{EVENT}} > \alpha_3 \times \overline{\text{EFF}}\right)\end{aligned} \tag{12}$$

where α_2 and α_3 are empirically set to 1.1 and 0.85. F_{EVENT} corresponds to the filtered events after event filtration. $\overline{\text{ECF}}$ and $\overline{\text{EFF}}$ indicate the average value of the event categorization feature and event filtration feature, respectively. Events which satisfy Equation 12 are referred as high-impact events while the rest of the events are put in low-impact class of events. At the end of the event categorization stage, we obtain high-impact and low-impact sets of events. Other events may remain present in both these classes because player clash, foul and injury are such events which may exist for a longer span or shorter span.

2.4 Edge and caption analysis

After obtaining the high-impact events, we analyse them using the edge pixel ratio and contents of the caption. Goal event is mostly followed by cheering in the audience view as well as gathering of players as well as goal post views which may give rise to edges in the frames. After the occurrence of goal events, every broadcaster displays the caption about the goal information. 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 as shown in Figure 7a. This caption almost stays for 4 to 8 s. After the completion of an event, the broadcaster displays the caption of team score information at the bottom part of the frame (image) as shown in Figure 7b. Generally, it is observed that the goal score caption is displayed within 55 s after the goal event. The presence of such views and detailed caption becomes a very important clue for the confirmation of the goal event. We do not consider initial frames up to 4 s (100 frames) of an event for EPR computation, and we continue to compute EPR for another 55 s (1,250 frames) even after the end of an event for the inclusion of the goal score caption.

We apply the following steps to carry out edge analysis of high-impact events:

1. Divide the image horizontally into three equal parts and choose the bottom part where the caption is located.
2. Convert the image in grey and apply the Sobel operator to detect horizontal edges.
3. Compute EPR.

EPR is computed using the formula mentioned in Equation 5.

The EPR of every frame of an event is computed, and then we compute the average value of the EPR of an event and also the average EPR of all high-impact events. Finally, we select such events whose EPR value is greater than the threshold. The threshold is empirically set to 0.97 times the

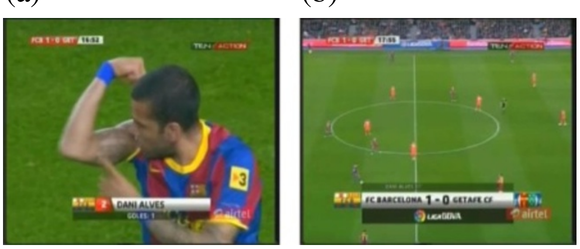
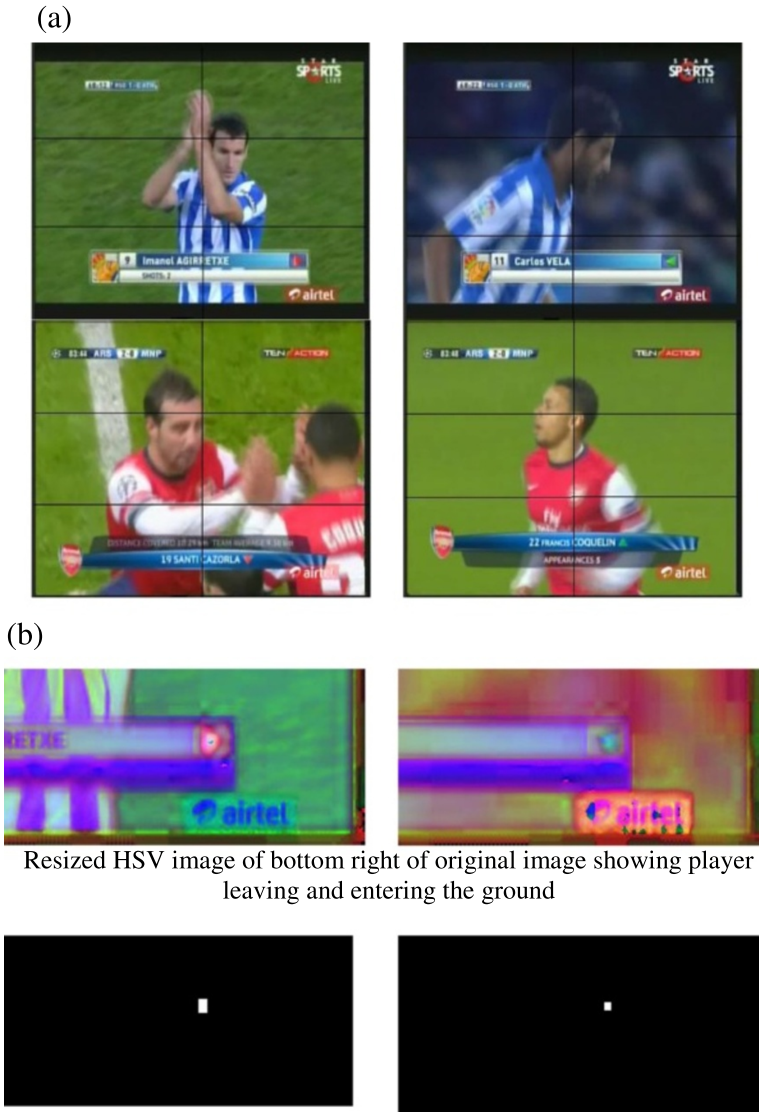


Figure 7 Goal caption displaying the player and team information. (a) Goal caption displaying the player's name. **(b)** Goal score caption displaying team names.

mean of the EPR of high-impact events. Events which have a higher EPR also include player exchange events. Player exchange events experience huge motion as the players are replaced on the ground. Many times, it involves transitions among far (ground), close-up and audience views similar to the goal event. 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 colour. These types of captions and the results of the below mentioned algorithm are shown in Figure 8a,b, respectively. Now, we analyse the high EPR events based on the nature of their caption. A brief algorithm has been described below to separate the player exchange events from the high EPR events. The algorithm process is much more similar to the yellow card detection algorithm.

1. Divide the image horizontally into three and vertically two equal parts and choose the bottom right part where the caption is located.
2. Resize the bottom right image by a factor of 2.
3. Obtain the HSV image and search for the triangle symbol made of red/green pixels in the HSV image.



Resized HSV image of bottom right of original image showing player leaving and entering the ground

Exactly hitting the Red and Green spot indicating Player Exchange Events

Figure 8 Player exchange events and detected symbols. (a) Various images of player exchange events and divided into six regions. **(b)** Detected symbols of player exchange events.

4. Obtain the connected components of the image which has an area within the specified range.
5. Keep those frames as player exchange frames if the minimum number of frames within that segment meets the defined criteria.
6. Find the distance between the red and the green spot of the selected images; if the distance is less than threshold, then declare a player exchange event.

We resize the image by a factor of 2 to have enough large area of red/green triangle for proper detection. The hue range for red has been set to more than 0.90, and for green, it has been set between 0.30 and 0.40, while saturation and value for both red and green colours are set more than 0.80 and 0.60, respectively, in step 3. This region is smaller; hence, we cannot exactly search for the triangular shape or symbol, but we only search for pixels which belong to the above specified range of hue, saturation and value. The threshold for the area of the symbol is set between 10 and 150 in step 4 after observing various sizes of the symbol of player exchange events. The minimum number of frames which are required to declare a player exchange event is set to 15 for both the red and green symbols. The threshold for the city block distance between the top left coordinates of the red and green spots is empirically set less than 20. Detected player exchange events are removed from the set of events which have high EPR, and we refer to the remaining set of events as probable goal events.

2.5 View classification

After edge analysis, two sets of probable goal and low-impact events. In order to appropriately classify or label these events, it is necessary to realize the temporal pattern of the frames of an event. To furnish this task, it is necessary to label every frame of the event of the video. This process is referred to as view classification. Since this process is entirely independent of event filtration and categorization, it can be applied in parallel. In order to carry out view classification, we extract visual features from the frames and classify them into one of the predefined views. Characteristics of different views are described below:

- *Far field view*: A far field view displays a global view of the game field. It is captured by a camera at a long distance. It is often used to show the play status, such as play position and long passes. In this view, the ratio of the field area to the whole image is high, and the size of players within the field is small.
- *Goal post view*: A goal post view displays a goal post area. It is shown when players are attempting to get a goal. If the goal post is captured from a long distance, the ratio of the field area is high; otherwise, the field ratio is medium to low. The goal post view is partially dominated by audience view.
- *Medium field view*: A medium view is a zoom-in view of a specific part of the field. It usually shows players and referees with the field as a background. In a medium view, the size of players in the playfield is bigger than that in a long view and the field ratio is in the medium range.
- *Close-up view*: An outfield view displays close-up of players, coach or players gathering with non-field background. It often focuses on the leading actor of current event. In this view, the field ratio is very low.
- *Audience view*: An outfield view displays the audience, as an indication of a break caused by highlights, such as an audience cheer view after a goal. In the audience view, the field ratio is extremely low and generally texture is dominant and complex.

The view classification system is shown in Figure 9. At the first level of classification, algorithm I is applied on all the frames of an event of the video for far field view and non-far field view classifications.

Algorithm I: field view detection

1. Convert the input frame from an RGB image into an HSV image.
2. Get the hue histogram of the image.
3. Define the hue range, which covers the different variations of the playfield's green colour, as a green window.
4. Compute the grass pixel ratio (GPR).
5. Apply the *K*-means algorithm on GPR to cluster frames into two clusters, one with high GPR values and the other with low GPR values.

The playfield usually has a distinct tone of green that may vary from stadium to stadium of different leagues of soccer. Matches that are played under floodlight exhibit different tones of green than sunlight. Even the shadow effect is also observed on the playfield many times under sunlight which also affects the intensity of green colour. So, hue range, which can cover different playfields' green colour, is carefully decided and identified as green range. The range of hue for the identification of various shades of green is set between 0.23 and 0.38 which we can refer to as a green window. We also involve the saturation and value components by setting them greater than 0.40. Due to varying green tones, the grass pixel ratio differs largely on various datasets; hence, it is not wise to set the threshold statically for the separation of far field and non-far field views. Instead, in step 5, we apply *k*-means to separately cluster these views. Algorithm I classifies

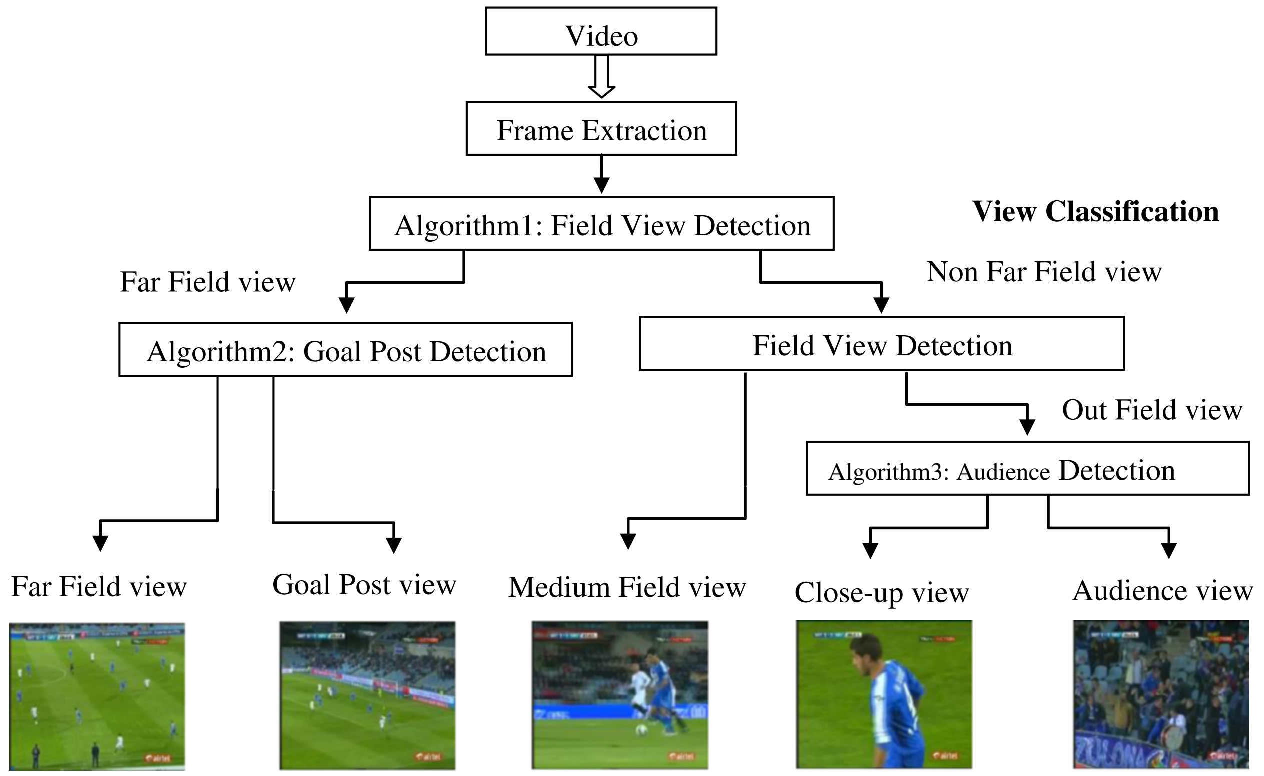


Figure 9 View classification system.

each frame in either far field view or non-far field view. The proposed goal post view detection method is mentioned below.

Algorithm II: goal post detection

1. Convert the input RGB image into a grey scale image.
2. Apply the Sobel edge detection operator on the grey image to detect vertical edges.
3. Erode the image with a vertical structuring element.
4. Apply a canny edge detection operator on the grey image to detect field lines near the goal post.
5. Apply Hough transformation.
6. If vertical parallel lines and parallel lines on the field are detected, then the frame belongs to a goal post view.

The Sobel edge detection operator is applied on the image to detect vertical lines. The resultant image exhibits vertical lines of the goal post as well as many other vertical lines, whose length is less than that of goal post lines. To remove such unimportant lines, erosion operation with a vertical structuring element is applied on the resultant image. We have used a vertical structuring element of length 5.

The output of edge detection operation is an image described by a set of pixels having vertical edges. This set of pixels rarely characterizes an edge completely because of noise and breaks in the edge. So, edge detection operation is followed by edge linking technique to assemble edge pixels into meaningful edges. We apply the Hough transform to detect linked vertical edges. If parallel vertical lines of the goal post and parallel field lines are detected, then we can conclude that the frame is having a goal post view. Figure 10c shows the detected two vertical poles of the goal post; however, these edges may be broken or noisy. Hence, results of the Hough transformation in Figure 10c are shown in Figure 10d. For the detection of parallel field lines near the goal post which are partially horizontal, the canny edge detection method is applied. Canny is a good candidate for thin as well as dull edges; we opt canny detection for these horizontal edges. Figure 10e depicts the existence of field lines near the goal post, and Figure 10f shows the result of the Hough transformation. Figure 11a,b,c,d,e,f shows the results of the goal post detection method of a left-oriented goal post.

2.6 Audience view detection

Classification of audience view and close-up view is based on finding EPR. Edge images generated using

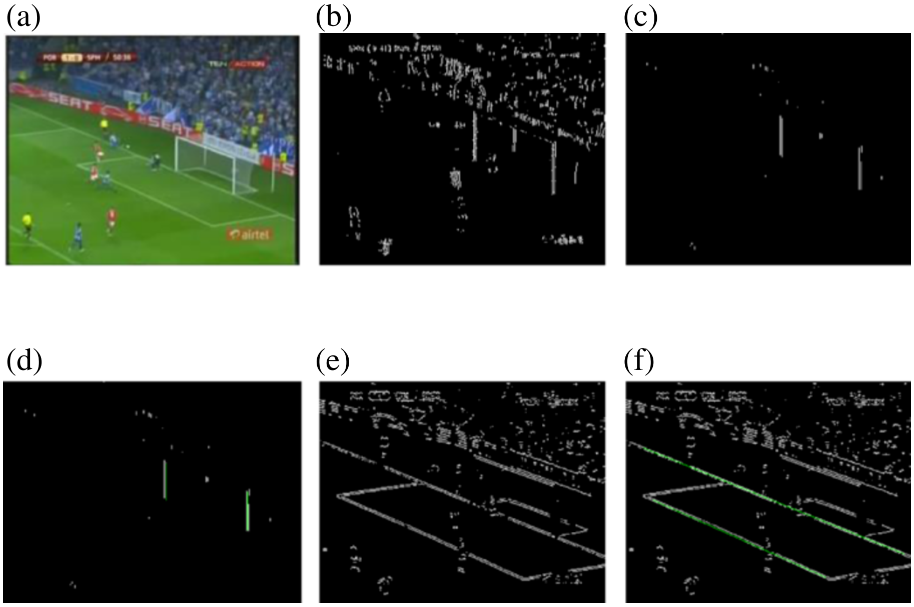


Figure 10 Goal post view (right-oriented) detection algorithm results. (a) Goal post view. **(b)** Vertical edge detection result of **(a)** using Sobel. **(c)** Erosion of image **(b)**. **(d)** Hougth transformation result of **(c)**. **(e)** Horizontal edge detection result of **(a)** using canny. **(f)** Hough transformation result of **(e)**.

canny edge detection are shown in Figure 12b,d along with their EPR values. The EPR value of audience view is quite higher than that of close-up view. EPR is statically set to 4.5 by experimenting on a large number of frames of different condition videos. The audience view detection algorithm has been described below.

Algorithm III: audience view detection

1. Convert the input RGB image into a grey image.
2. Convert the grey image into a binary image using canny edge detection operator.
3. Compute EPR as shown in Equation 5.

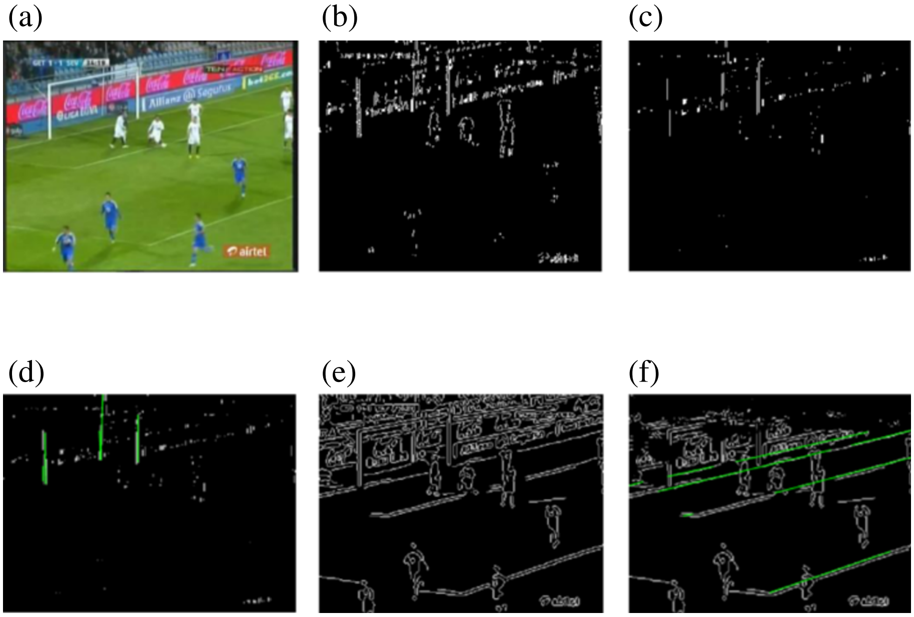


Figure 11 Goal post view (left-oriented) detection algorithm results. (a) Goal post view. **(b)** Vertical edge detection result of **(a)** using Sobel. **(c)** Erosion of image **(b)**. **(d)** Hougth transformation result of **(c)**. **(e)** Horizontal edge detection result of **(a)** using canny. **(f)** Hough transformation result of **(e)**.

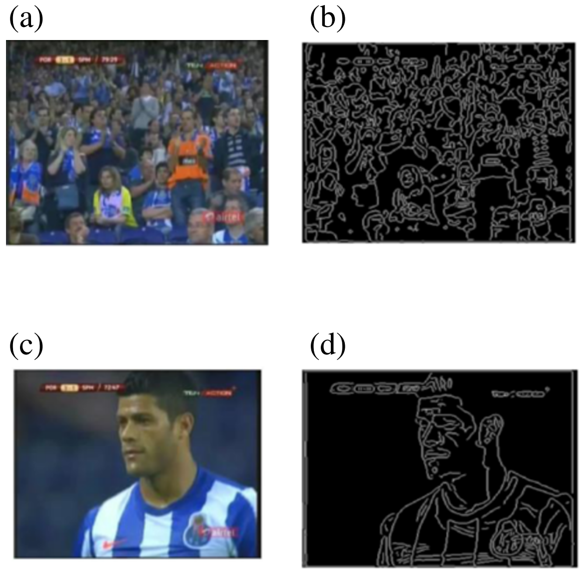


Figure 12 Audience and close-up views. (a) Audience view: frame no. 42723, video 10. **(b)** Edge image of left EPR: 6.18. **(c)** Close-up view, frame no. 33718, video 10. **(d)** Edge image of left EPR: 2.29.

4. Define the edge pixel threshold (EP_{th}) for audience view classification.
5. If EPR > EP_{th}, then
   i. The frame is classified as audience view.
6. Else
   i. The frame is classified as close-up view.

3 Event classification using HMM

At the end of the view classification phase, we end up with classification of every frame of an event in one of the above-mentioned views. To recognize certain interesting events from sports videos, temporal transition patterns among the frames (views) within the event can be utilized. To formulate the pattern of an event, it is necessary to map low-level features into high-level semantic. For example, to detect a goal scoring event, there are temporal patterns such as more transitions that happen among goal post, close-up and audience views because every goal event is followed by gathering of players and cheers in the audience. If these patterns can be recognized for the event sequence, the corresponding events can be identified. To boost up the efficiency of the event classification, it takes low-impact and high-impact classes of events as inputs which are produced by the event categorization phase. Since card events are already detected using the yellow card event detection algorithm, card event segments are removed from the high- or low-impact class of events. The event classification system proposed here uses the hidden Markov model to classify goal event, goal attack and other events in a soccer video. High-impact events are classified into either a goal event or other events using a trained goal event model and other event models. Similarly, low-impact events are classified into goal attack or other events using a trained goal attack event model and other event models. HMM is described in following section.

3.1 Hidden Markov model

HMM is a statistical model for dealing with hidden states and observations. A video semantic event forms a Markov process, so the HMM is adopted as a powerful tool for video content analysis.

A HMM is defined by

- a set of states, Q
- a set of transitions, where transition probability $a_{\mathrm{kl}} = P(\pi_{\mathrm{i}} = l \mid \pi_{\mathrm{i}\ -1} = k)$ is the probability of transitioning from state k to state l for $k, l \in Q$
- an emission probability, $e_k(b) = P(x_i = b \mid \pi_i = k)$, for each state, k, and each symbol, b, where $e_k(b)$ is the probability of seeing symbol b in state k. The sum of all emission probabilities at a given state must equal 1, that is, $\Sigma_b e_k = 1$ for each state, k.

There are basically three problems which can be solved using HMM.

Evaluation: Given an observation sequence x, and model parameters, determine the probability, $P(x)$, of obtaining sequence x in the model. The solution to this is the forward-backward algorithm.
Decoding: Given an observation sequence and model parameters, determine the corresponding optimal state sequence. This problem solution is the Viterbi algorithm.
Learning: Given a model and a set of training sequences, find the model parameters (transition and emission probabilities) that explain the training sequences with relatively high probability. The Baum-Welch algorithm is used for this purpose.

The steps of the event detection process are briefly described below.

3.2 Event detection using HMM

1. Each event corresponds to one model. In classification, we generated three models corresponding to goal, goal attack/corner and other events. Other event classes include events like foul, free pass, player exchange etc. which do not produce much impact on the nature of the game.
2. Given an observation sequence $O = O_1, O_2, \ldots, O_{\mathrm{n}}$ produced from a shot sequence, the probability $P(O|\lambda_i)$ relative to each model λ_i is calculated by using

the Viterbi algorithm. Thus, this sequence belongs to event$_j$ where $j = \arg\max_{i=1\ldots N} P(O/\lambda_i)$.

To achieve a higher recognition rate of the events correctly, at the end of the event categorization stage, we successfully separate the low-impact and high-impact events. Training the model also poses several questions like with how many events should we train the model to obtain highly accurate recognition. Normally, a goal event consists of transitions among the far, goal post, close-up and audience views in which close-up and audience views are dominating. A general observation infers that every goal is followed by cheering in the audience; however, in several goal events, the camera is not much more focused to the audience and their cheering. Prior to the goal event, occurrence of penalty kick types of events can also change the general pattern of the goal event. This situation applies to every rest event of soccer. Models of goal and other events are applied to the high-impact class while goal attack and other event models are applied to the low-impact class.

4 Simulation results and discussions

We have experimented with soccer videos of total length almost 6 h. We have conducted experiments on 13 video datasets from seven well-known soccer leagues like Barclays Premier League, La Liga, Serie A Premier League, FIFA EURO CUP, Europa, England 2 and Champions League. All these videos possess varying ground and illumination conditions, e.g. daylight, floodlight, rainy as well as shadow. Video datasets have a 352 × 288 resolution. Video dataset information is shown in Table 1 along with the video illumination condition. Date information of the match is shown in DD/MM/YYYY format. Varying ground and illumination conditions are clearly depicted in Figure 13 where we have depicted the far view images of different videos used in the experiments. We can easily observe that every far view is highly different than the far view of other videos as this is natural because videos belong to different leagues and different illumination conditions. Datasets used in the experiments are available on the web at http://dspinnovations.blogspot.in/. We aim to propose a novel framework for a large number of various leagues and every type of illumination conditions. Table 2 reflects the average variance of the red, green and blue components of far-view-classified images of videos. Each video exhibits a large amount of difference in the average variance of the green component with the other one. Due to this fact, it is also observed that every video also differs largely in the number of far views which are classified by view classification. Soccer is a highly eventful game in which events like card, penalty kick, corner, foul, throw in, off side, goal attempts, goal etc. usually happen frequently, but card, goal and goal attack/corner types of events attract user's attention. We do not consider the repeated segments (replay) of an event because after the occurrence of the event replay is played, we only classify the original event.

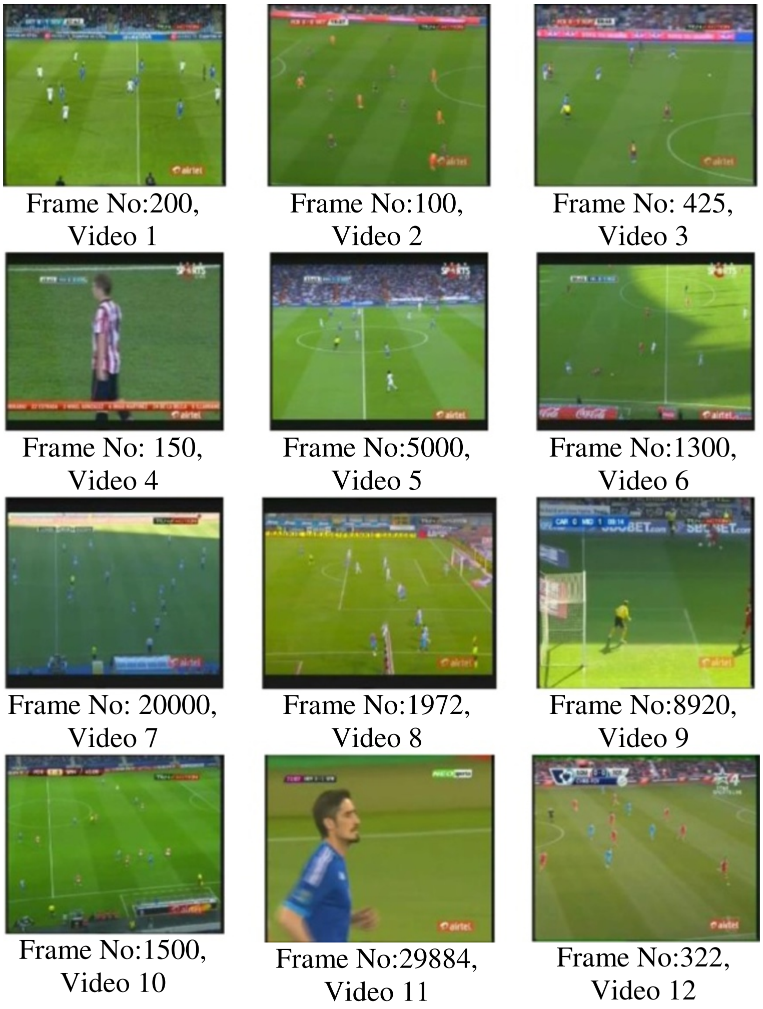
Experimental results are evaluated using standard parameters: precision and recall. Precision quantifies what proportion of the detected events is correct while recall quantifies what proportion of the correct events is detected. If we denote D the events correctly detected by the algorithm, D_m the 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:

$$\mathrm{Recall} = \frac{D}{D + D_\mathrm{m}}$$

$$\mathrm{Precision} = \frac{D}{D + D_\mathrm{f}}$$

Table 1 Soccer video information

Serial number	Match information (team name, league name, date information, condition)	Duration in minutes
1	Getafe vs Sevilla, 1st half, April 2012, La Liga 2012 16/4/2012, floodlight	16:00
2	FC Barcelona vs Getafe, 1st half, March 2011, La Liga 2011, 19/3/2011, floodlight	10:00
3	FC Barcelona vs Almeria, 2nd half, La Liga 2011, 19/3/2011, floodlight	14:05
4	Real Sociadad vs Athletic Club, 2nd half, La Liga, 2011 29/9/2012, floodlight	33:00
5	Real Madrid vs Deportivo, 1st half, La Liga, 2011 30/9/2012, floodlight	22:00
6	Celta Vigo vs RCD Mallorca, 2nd half, La Liga, 2011 18/11/2012, daylight, shadow	20:00
7	Udinese vs Bologna, Serie A, 2nd half, 2/10/2011, daylight, shadow	16:10
8	Catania vs Intermilan, Serie A, 1st half, 15/10/2011, floodlight, rainy	16:00
9	Cardiff vs Middlesbrough, 1st half, England 2, 2/5/2011, daylight, shadow	35:00
10	Porto-Spartak Moscow, 2nd half, Europa League 7/4/2011, floodlight	40:00
11	Germany vs Greece, 2nd half, FIFA WC 2012, 21/6/2012, floodlight	38:00
12	Southampton vs Spurs, 1st half, Barclays Premier League, 22/12/2013, daylight, shadow	48:00
13	Arsenal vs Montpellier, 2nd half, Champions League, 21/11/2012, floodlight	46:00



Frame No:200, Video 1

Frame No:100, Video 2

Frame No: 425, Video 3

Frame No: 150, Video 4

Frame No:5000, Video 5

Frame No:1300, Video 6

Frame No: 20000, Video 7

Frame No:1972, Video 8

Frame No:8920, Video 9

Frame No:1500, Video 10

Frame No:29884, Video 11

Frame No:322, Video 12

Figure 13 Various far views depicting largely varying ground conditions.

Table 2 Average variance of far views of videos

Dataset	Number of far views detected	Average variance of RGB of far views		
		R	G	B
La Liga 1	8,745	926.50	1,188	1,969
La Liga 2	9,533	963.99	813.23	1,213
La Liga 3	14,653	810.81	765.11	1,232
La Liga 4	9,374	1,711	2,063	2,148
La Liga 5	18,968	1,661	2,774	2,018
La Liga 6	13,333	1,325	2,059	1,602
Serie A 1	1,949	2,998	3,125	2,132
Serie A 2	365	2,267	2,773	2,946
England 2	1,477	2,088	1,741	1,713
Europa	37,831	808.22	1,245	1,254
EURO CUP	182	3,379	2,388	3,473
Barclays	345	184,637	195,797	299,013
Champions	5,981	88,651	120,522	159,864

Table 3 shows the performance of the card event detection algorithm. Results are very encouraging and achievement of 100% recall also sustains excellent precision of more than 95% in spite of having different types of cards like tilted, smaller or larger in size and varying shades of yellow. Our framework does not have any constraints, for example, in [32], for card event detection, the approach relies on the detection of the referee wearing a black t-shirt. Also, very limited types of soccer leagues have been experimented in [11,32] which possess almost similar types of ground and illumination conditions. The proposed yellow card event detection algorithm is found generic as it overcomes all such mentioned limitations. Our proposed algorithm succeeds and does not detect the yellow card event in the soccer dataset which does not contain the card event. Table 4 shows the number of events (shots) detected in every video and also the number of

Table 3 Performance of the yellow card detection algorithm

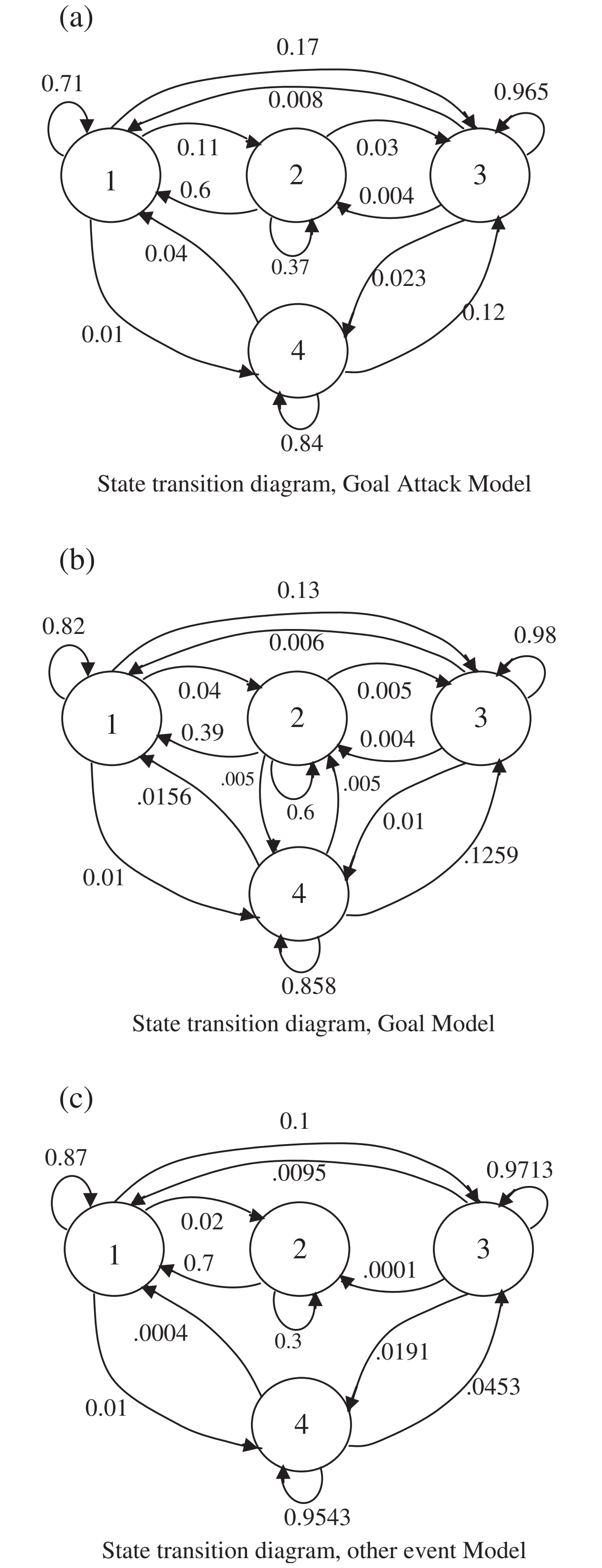
League name	Total yellow cards	Detected	Precision (%)	Recall (%)
La Liga 1	1	1	100	100
La Liga 2	1	1	100	100
La Liga 3	2	2	100	100
La Liga 4	2	2	100	100
La Liga 5	2	2	100	100
La Liga 6	1	1	100	100
Serie A 1	1	1	100	100
Serie A 2	0	0	100	100
England 2	3	3	100	100
Europa	3	4	75	100
EURO CUP	1	1	100	100
Barclays	1	1	100	100
Champions	3	3	100	100
Total	21	22	95.4	100

low- and high-impact events detected after event filtration and event categorization.

Every goal event has much more similarity to the other goal events while goal attack events may differ slightly because goal attack can happen from the frontal side of the goal post or from the corner (corner event). There is no specific criterion about the number of samples (events) required to accurately train the system, even overtraining can also result in poor classification accuracy. The corner/goal attack event has been trained using seven event shots

Table 4 Number of events detected as low/high-impact events after various stages

Soccer league	Number of events detected after threshold *T* (Equation 5)	Number of events after event filtration	Number of events after event categorization	
			Low-impact	High-impact
La Liga 1	12	7	6	1
La Liga 2	11	6	4	2
La Liga 3	22	11	8	3
La Liga 4	44	22	17	5
La Liga 5	24	11	9	2
La Liga 6	30	11	7	4
Serie A 1	23	11	6	5
Serie A 2	22	13	10	3
England 2	43	27	19	8
Europa	54	32	22	10
EURO CUP	36	23	18	5
Barclays	59	36	30	6
Champions	47	19	15	4



State transition diagram, Goal Attack Model

State transition diagram, Goal Model

State transition diagram, other event Model

Figure 14 State transition diagrams. (a) Goal attack model. **(b)** Goal model. **(c)** Other event model.

Table 5 Goal event classification results

Soccer league	Total goals	Detected	Correct	False	Missed	Precision (%)	Recall (%)
La Liga 1	1	1	1	0	0	100	100
La Liga 2	1	1	1	0	0	100	100
La Liga 3	2	2	2	0	0	100	100
La Liga 4	2	2	2	0	0	100	100
La Liga 5	2	2	2	0	0	100	100
La Liga 6	1	1	1	0	0	100	100
Serie A 1	1	1	1	0	0	100	100
Serie A 2	1	2	1	1	0	50	100
England 2	4	4	2	2	2	50	50
Europa	5	6	5	1	0	83.33	100
EURO CUP	6	6	5	1	1	83.33	83.33
Barclays	3	3	2	1	0	66.67	100
Champions	2	1	1	0	1	100	50
Total	31	32	26	6	4	81.25	86.7

while goal and other event models are trained using eight and ten event shots, respectively. As a result of training, the generated state transition diagrams to classify goal attack, goal and other events are shown in Figure 14. Numbers 1, 2, 3 and 4 indicate far view, goal post view, close-up/mid-field view and audience view, respectively. Since goal events are lengthy events, they exhibit interrelationships between every state. Other events include foul, throw in and player injury types of events. It is natural to realize that transitions between goal post to audience and between goal post to close-up are not observed in the other event models. The goal attack event model exhibits high transition probabilities between goal post view and close-up view compared to other event models as the event has the dominance of goal post view and its associated transitions.

Results of the event classification stage are shown in Tables 5, 6 and 7. For the goal event, we achieve a very good recall rate of 86.7% and precision of 81.25% in spite of various leagues. Several genuine goal attacks are also misclassified as goal events because many times goal attacks are followed by a player's disappointment and disgust of the audience which also create similar kinds of transition patterns among states like goal events. There are also goal events in which the camera is not focused on the audience or even does not follow more celebration by the players. Such events are misclassified due to the change in their pattern. We achieve a very

Table 6 Goal attack/corner event classification results

Video	Total	Detected	Correct	False	Missed	Precision (%)	Recall (%)
La Liga 1	5	5	5	0	0	100	100
La Liga 2	1	3	1	2	0	33.33	100
La Liga 3	4	5	4	1	0	80	100
La Liga 4	6	4	4	0	2	100	66.67
La Liga 5	3	5	3	2	0	60	100
La Liga 6	5	6	5	1	0	83.33	100
Serie A 1	3	4	3	0	0	100	100
Serie A 2	2	3	1	2	1	33.3	50
England 2	9	10	8	2	1	80	88.89
Europa	13	11	11	0	2	100	84.6
EURO CUP	8	10	6	4	2	60	75
Barclays	9	17	8	9	1	47.05882	88.89
Champions	8	11	8	3	0	72.72727	100
Total	76	94	67	26	9	72.04	88.15

Table 7 Other (throw in, offside, injury etc.) event classification results

Video	Total	Detected	Correct	False	Missed	Precision (%)	Recall (%)
La Liga 1	0	0	0	0	0	100	100
La Liga 2	3	0	0	0	3	100	0
La Liga 3	2	1	1	0	1	100	50
La Liga 4	11	13	11	2	0	84.6	100
La Liga 5	2	0	0	0	2	0	0
La Liga 6	6	4	4	0	2	100	66.7
Serie A 1	3	2	2	0	1	100	66.7
Serie A 2	8	5	4	1	4	80	50
England 2	7	4	2	2	5	50	28.5
Europa	7	8	7	1	1	87.5	87.5
EURO CUP	9	6	5	1	5	83.3	50
Barclays	23	10	9	1	14	90	39.1
Champions	5	0	0	0	5	0	0
Total	86	53	45	8	43	84.9	51.1

good precision for other events of 84.9% and a recall of 51%. Goal attack is also well recognized with a high recall of 88.1% as well as a quite good precision of 72%. Many times, other types of events which happen near goal posts are wrongly classified as goal attack due to the presence of the goal post in many views, so this reduces the precision of the goal attack event and recall of other events. Other events themselves consist of many events like foul and injury where each could even have different transition patterns. In spite of this fact, the framework achieves very good precision. There is no labelled and accurate dataset that is available for soccer videos. Hence, comparison cannot be considered fair, but we have compared our results with the results of [32]. Comparative performance is shown in Figure 15. The proposed approach proves its soundness as it outperforms in precision. Also, in [32], the authors have experimented with only two types of leagues while we have used seven various types of leagues.

We have analysed our proposed framework with different conditions like floodlight, daylight, shadow and dim illumination, and rainy condition. These conditions lead to major variation in illumination, and in this context, it is our major achievement to have a framework which successfully works in the presence of such conditions. However, we obtain little less recall in goal and corner/goal attack events. There are two clear reasons for this: 1) every league has its intrinsic characteristics of movement of cameras during the entire match as well as after the occurrence of an event. This pattern slightly differs in other leagues, e.g. a corner event is shown with more close-up views in one league while in another same event with more far views in another league. Hence, it is difficult to obtain precise view classification

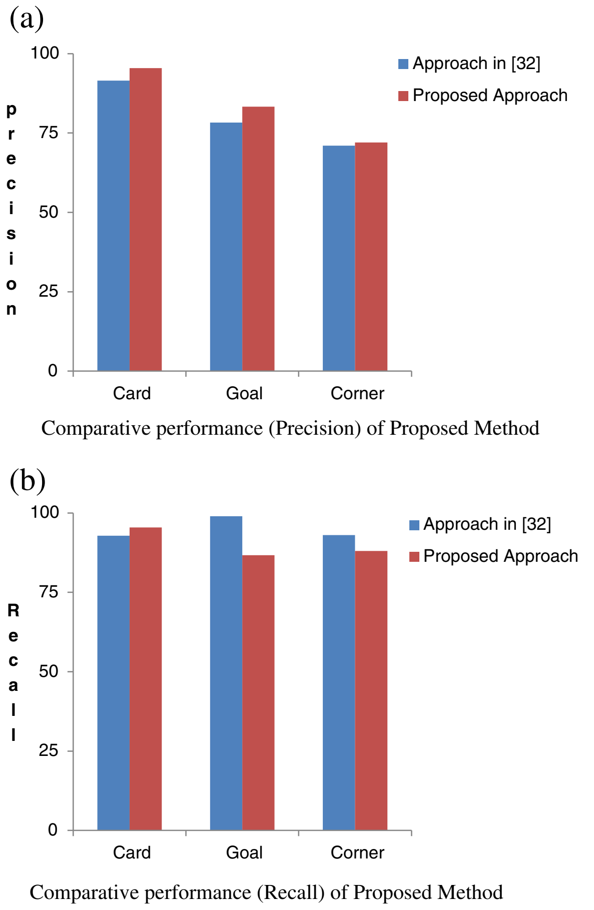


Figure 15 Comparative performance. (a) Precision of the proposed method. **(b)** Recall of the proposed method.

over different leagues. 2) There is no labelled dataset in soccer, and there is no ground truth available for the beginning and ending marks of various events within the match, so one cannot say exactly which are the beginning and ending frames of an event. We carry out the task of event segmentation automatically by our proposed method, but no method can be accurate to mark exactly this span of an event. It is important to note that our framework is not tuned to any specific broadcast video. Our framework needs only an input video and each video will be converted into a set of images for further processing. The input videos to our framework are broadcast ones, and hence, it is worth to mention that there is no control over the type of input video. Our framework does not have any constraints on frames per second.

5 Conclusions

In this paper, a novel, effective, robust fully automatic framework is proposed to detect and classify the important events of the soccer videos of various leagues under various illumination conditions. The proposed framework effectively uses the optical flow to demarcate the event. The proposed algorithm for card event detection achieves very high accuracy and found invariant to scale, tilt and varying yellow shades of the card. In order to improve the classification accuracy of the system, event filtration and event categorization processes are applied. These processes do not only contribute in improving the efficiency but also make the framework robust to the varying condition datasets. Event classification is successfully carried out by hidden Markov models after obtaining low-impact and high-impact classes of events. HMM can be easily adopted because every event has its own transition pattern. The generated results clearly reflect the efficiency and efficacy of the framework with different kinds of leagues. It is also important to note that our approach is fully automatic in the sense that our framework is able to classify all the detected events with high precision. Experimentation and investigations are still under progress to develop a framework which can be applicable to a large number of various soccer leagues.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]U V Patel College of Engineering, Ganpat University, Kherva 384012, Mehsana, India. [2]Sardar Vallabhbhai National Institute of Technology, Surat 395007, India.

Received: 14 April 2014 Accepted: 20 October 2014
Published: 3 November 2014

References

1. L Ying, Z Tong, T Daniel, *An Overview of Video Abstraction Techniques*. Technical Report. HP-2001-191 (HP Laboratories, Palo Alto, 2001)
2. M Roach, J Mason, L-Q Xu, F Stentiford, Recent trends in video analysis: a taxonomy of video classification problems, in *Proceedings of the 6th International Conference on Internet and Multimedia Systems and Applications (IASTED)* (Hawaii, 2002), pp. 348–353
3. S Carrato, I Koprinska, Temporal video segmentation: a survey. Signal Process. Image Commun. **16**(5), 477–500 (2001)
4. AG Money, A Harry, Video summarization: a conceptual framework and survey of the state of the art. J Visual. Commun. Image. Represent **19**(2), 121–143 (2008). doi:10.1016/j.jvcir.2007.04.002
5. G Yue, D Hai, Shot based similarity measure for content based video summarization, in *Proceedings of 15th IEEE International Conference on Image Processing (ICIP)* (San Diego, 2008), pp. 2512–2515
6. H Chen, L Zeng, Indexing and matching of video shots based on motion and color analysis, in *Proceedings of 9th IEEE International Conference on Control, Automation, Robotics and Vision (ICARCV)* (Singapore, 2006), pp. 1–6
7. S Wei, Jiang, A novel algorithm for video retrieval using video metadata information, in *Proceedings of IEEE International Workshop on Education Technology and Computer Science (ETCS)*, vol. 2 (Wuhan, 2009), pp. 1059–1062
8. W Wolf, Key frame selection by motion analysis, in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (Atlanta, 1996), pp. 1228–1231
9. B Lucas, T Kanade, An iterative image registration technique with an application to stereo vision, in *Proceedings of Seventh International Joint Conference on Artificial Intelligence (IJCAI)* (Vancouver, 1981), pp. 674–679
10. S Ling, J Ling, Motion histogram analysis based key frame extraction for human activity representation, in *Proceedings of Canadian Conference on Computer and Robot Vision (CRV)* (Kelowna, 2009), pp. 88–92
11. B Li, MI Sezan, Event detection and summarization in American football broadcast video. Proc. SPIE **4676**, 202–213 (2002)
12. R Leonardi, P Migliorati, Semantic indexing of multimedia documents. Proc IEEE. Multimedia **9**, 44–51 (2002)
13. J Liu, X Tong, W Li, T Wang, Y Zhang, H Wang, Automatic player detection, labeling and tracking in broadcast soccer video, in *Proceedings of British Machine Vision Conference (BMVC)* (Warwick, 2007), pp. 1–10
14. X Lexing, SF Chang, A Divakaran, S Huifang, Structure analysis of soccer video with hidden Markov models, in *Proceedings of International Conference Acoustics, Speech and Signal Processing (ICASSP)* (Orlando, 2002), pp. 4096–4099
15. W Zhou, A Vellaikal, C-C.J. Kuo, Rule-based video classification system for basketball video indexing, in *Proceedings of ACM Multimedia Conference* (Los Angeles, 2000), pp. 213–216
16. D Zhang, S-F Chang, Event detection in baseball video using superimposed caption recognition, in *Proceedings of 10th ACM of International Conference Multimedia* (Juan Les Pins, 2002), pp. 315–318
17. V Kiani, HR Pourreza, Flexible soccer video summarization in compressed domain, in *Proceedings of International Conference on Computer and Knowledge Engineering (ICCKE)* (Mashhad, 2013), pp. 213–218
18. DW Tjondronegoro, YP Chen, Knowledge discounted event detection in sport video. IEEE Sys. Man. Cybern Part A: Syst. Hum **40**(5), 1009–1024 (2010). doi:10.1109/TSMCA.2010.2046729
19. H Chung-Lin, C Chih-Yu, Video summarization using hidden Markov model, in *Proceedings of International Conference on Information Technology: Coding and Computing (ITCC)* (Las Vegas, 2001), pp. 473–477
20. W Jinjun, X Changsheng, C Engsiong, T Qi, Sports highlight detection from keyword sequences using HMM, in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)* (Taipei, 2004), pp. 599–602
21. J Assfalg, M Bertini, A Del Bimbo, W Nunziati, P Pala, Soccer highlights detection and recognition using HMMs, in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)* (Lausanne, 2002), pp. 825–828
22. G Xu, YF Ma, HJ Zhang, SQ Yang, An HMM-based framework for video semantic analysis. IEEE Trans. Circuits. Syst. Video. Technol **15**(11), 1422–1433 (2005). doi:10.1109/TCSVT.2005.856903
23. C Wu, Y-F Ma, H-J Zhang, Y-Z Zhong, Events recognition by semantic inference for sports video, in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, vol. 1 (Lausanne, 2002), pp. 805–808
24. DA Sadlier, N O'Connor, S Marlow, N Murphy, A combined audio-visual contribution to event detection in field sports broadcast video, in *Proceedings of IEEE Symposium on Signal Processing and Information Technology (ISSPIT)* (Darmstadt, 2003), pp. 552–555
25. M Petkovic, V Mihajlovic, W Jonker, S Kajan, Multi-model extraction of highlights from formula 1 programs, in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, vol. 1 (Lausanne, 2002), pp. 817–820

26. V Mihajlovic, M Pekovic, *Dynamic Bayesian Networks: A State of the Art* (CS Dept. Univ. Twente, Enschede, 2001)
27. S Alipour, P Oskouie, AM Moghadam, Bayesian belief based tactic analysis of attack events in broadcast soccer video, in *Proceedings of International Conference on Informatics, Electronics and Vision (ICIEV)* (Dhaka, 2012), pp. 612–617
28. Ren, Z Yuesheng, A Video Summarization approach based on machine learning, in *Proceedings of International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP)* (Harbin, 2008), pp. 450–453
29. J Basak, V Luthra, S Chaudhury, Video summarization with supervised learning, in *Proceedings of 19th International Conference on Pattern Recognition (ICPR)* (Tampa, 2008), pp. 1–4
30. K Ren, WAC Fernando, J Calic, Optimising video summaries using unsupervised clustering, in *Proceedings of 50th International Symposium (ELMAR)*, vol. 2 (Zadar, 2008), pp. 451–454
31. MS Hosseini, AM Moghadam, An adaptive neuro-fuzzy approach for semantic analysis of broadcast soccer video, in *Proceedings of International Conference on Development and Learning and Epigenetic Robotics (ICDL)* (San Diego, 2013), pp. 1–6
32. C Huang, H Shih, CY Chao, Semantic analysis of soccer video using dynamic Bayesian network. IEEE Trans Multimedia **8**(4), 749–759 (2006). doi:10.1109/TMM.2006.876289

doi:10.1186/1687-5281-2014-49
Cite this article as: Pandya and Zaveri: **A novel framework for semantic analysis of an illumination-variant soccer video.** *EURASIP Journal on Image and Video Processing* 2014 **2014**:49.