Artif Intell Rev (2014) 42:173–210
DOI 10.1007/s10462-012-9332-4

# Multimodal feature extraction and fusion for semantic mining of soccer video: a survey

**Payam Oskouie · Sara Alipour · Amir-Masoud Eftekhari-Moghadam**

Published online: 11 March 2012


**Abstract** This paper presents a classified review of soccer video analysis works. The existing approaches in the aspects of highlight event detection, video summarization and retrieval based on video stream, ball and player tracking for provision of match statistics, technical and tactical analysis and application of different sources in soccer video analysis have been surveyed. In addition, some major existing commercial softwares developed for video analysis are introduced and compared. With regard to the existing challenge for automatic and realtime provision of video analysis, different computer vision approaches are discussed and compared. Audio, video and text feature extraction methods have been investigated and the future trends for improvement of the reviewed systems have been introduced in terms of response time optimization, increase of precision and eliminating the need of human intervention for video analysis.

**Keywords** Soccer video analysis · Event detection · Video summarization · Field object tracking · Semantic mining

## 1 Introduction

According to the ever increasing advancement in multimedia technologies of video broadcasting and internet, large amount of digital videos from different sources and domains are accessible for users. These videos are provided and used in frequent qualities and volumes by the means of digital devices. In another hand, a large amount of these videos are dedicated

P. Oskouie (✉) · S. Alipour · A.-M. Eftekhari-Moghadam
Department of Electrical, Computer & IT Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran
e-mail: p.oskouie@qiau.ac.ir

S. Alipour
e-mail: s.alipour@qiau.ac.ir

A.-M. Eftekhari-Moghadam
e-mail: eftekhari@qiau.ac.ir

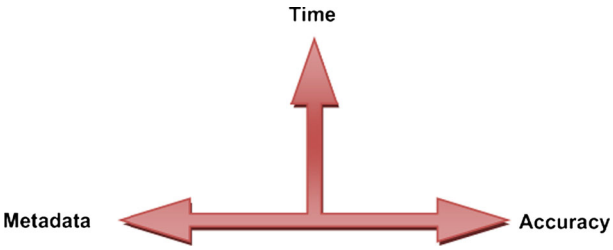


**Fig. 1** Effective parameters for video analysis performance

to soccer which is one of the most popular team sports in the world. As a result, knowledge acquisition in terms of annotation, indexing, semantic analysis, automatic event detection, match summarization, technical and tactical analysis, match statistic provision and similar applications have captures attention of researchers in different domains of image processing, video processing and video semantic analysis in the last decades. Semantic concepts extraction among these videos has its special difficulties. Obstacles which are faced by feature extraction from videos are challenging for the current approaches but can be addressed using available solutions for this domain (Gonzales and Woods 2008). Some problems such as low quality of videos, different lighting conditions, camera limited view, objects occluded by another, rapid occurrence of events, small size of objects in the video and some more similar issues have lead in various innovations and investigations for optimization of proposed approaches and algorithms.

All of the soccer analysis systems are designed based on three parameters of time, usage of metadata and accuracy (robustness). These systems are classified into two groups of offline and online or realtime from response time point of view. Online systems are those using complex and fast algorithms in which all processes are performed in realtime mode or by latency near to some seconds. These systems are applied in fast detection of important events such as goal, offside, booking as well as ball and players location. Offline systems have less complexity and more analytic view and their main goal is the important analysis which is obtained after the match. Team and player tactic analysis systems and all summarizations systems lie in this category. Regarding the dependence of the systems on metadata, they are categorized into dependent and independent. Metadata is information such as game log (organizable textual data which is created online by broadcasting channels and websites and reports the important events of the match for the users) or information such as jersey color of players, referee, etc. which is delivered to the system to make the process procedure fast and enhanced. The next effective parameter is the system accuracy in detecting the soccer events. Reviewing the works in literature shows that independent systems are less accurate in comparison with systems depending on metadata. In addition, providing high precision and decreasing error rate requires complex and time consuming processes that results the systems to work in offline mode. The effective parameters in complexity of the soccer analysis systems are displayed in Fig. 1.

The major components of a soccer video analysis system are according to Fig. 2. The input of system consists of video and attached metadata and video consists of images and audio itself. The metadata contains additional information of the video which assists the system for video semantic analysis. System inputs are processed in a component called “events and concepts detection”. In fact this component is the most significant part of video semantic analysis system which recognized and detects the events and concepts of the soccer match. In this component first some low level features are extracted from video. These features can be

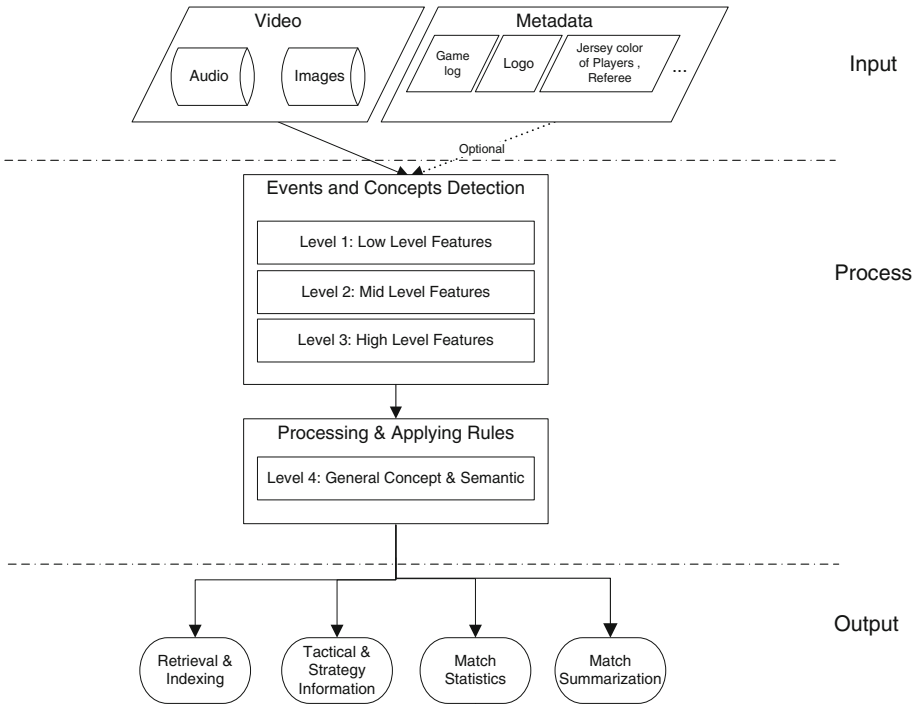


**Fig. 2** General scheme of soccer video analysis system

image color and texture, motion vectors, image segments, shot boundaries and audio. In the next step the middle level features such as ball, trajectories, players and referee are extracted from low level features (De Sousa J'unior et al. 2011). Mid-level features can be important objects of the video, closed captions, types of shots, scene and player movements. Afterwards high level features are detected which include the events and embedded concepts of the video segments. In general there are sixteen recognizable events available in a soccer video which are goal, penalty, offside, corner, booking foul, out, dribble, pass, head, game restart, player substitution, long pass, ball catching, ball possession and shot on goal. After video events and concepts detection, time stamp and other necessary information of the events are delivered to another component called "processing and applying rules" in which the events are processed and general concept and semantic of the video will be prepared. For instance, perception of team strategies can be obtained from player movements and team attacks-defenses as a general video concept.

The "events and concepts detection" component may be designed to work manually, semi-automatic and automatic. This issue is in direct relation with necessary amount of calculations (calculation complexity) and is in diverse relation with the precision of the system.

There are four types of output considerable for a soccer video analysis system which can vary based on the system application. The first type is a video clip which consists of the match summary. Usually users with different tastes are willing to view only the important and interesting events of a video and ignore the remaining parts due to time limitation. Summarization can be provided in different types based on user preferences. For instance, the output video can only consist of the match goals or be configured to display more highlight

events such as corners and shots on target. The second type of output includes statistical information about occurred events such as the number of goals, red and yellow cards, shots on target and ball possession. Provision of tactical information which is related to team and player strategies is considered as the third output of soccer video analysis system and is so demanding for soccer analysts and coaches. The last but not the least is soccer video indexing and retrieval. Content-based and semantic-based retrieval are two major approaches of video retrieval in the literature. In content-based video retrieval, the user provides an image or a small clip as a query and every video containing the similar frames and images to the query is extracted. Semantic-based retrieval delivers a higher level of output. In this approach, the semantic and essence lying in the stored videos is extracted automatically and then the related videos are retrieved based on the semantic of the query provided by the user.

This paper is organized as follows. In Sect. 2, detection of soccer highlight events is discussed. Video summarization and retrieval based on video stream features is reviewed in Sect. 3, the next section summarizes the ball and player tracking solutions and the works related to tactical analysis while different sources for video analysis are introduced in Sect. 5. In the last section, we discuss the future trends and challenges of video semantic mining for improvement of soccer video analysis and the commercial softwares in this relation have been compared.

## 2 Detection of soccer highlight events

The main goal of the most of soccer video analysis systems is automatic detection of the events which occur during a match and is important to soccer spectators, coaches, referees and in general anyone who deals with soccer issues. Regarding to the existing semantic gap between the reality of the events and the concept that a computer algorithm can understand and perceive, the researchers are trying to present more powerful and precise algorithms to minimize this semantic gap and provide a soccer analysis that is close to the real concepts as much as possible.

Everything that occurs in a soccer match and has a special meaning to an observer is called an event in the science of video semantic analysis. Each highlight scene includes some events which occur sequentially and this co-occurrence is interesting for the audience, such as the goal event which consists of ball entering the goal post and close-up of the players while gathering together and cheering (Kolekar et al. 2009a).

In general the soccer events can be categorized in two groups: the primary and secondary events. Events such as goal, penalty, booking, shot on target and offside situation which threatens goals of the teams are more effective in the result of the match comparing to the other events, therefore they are more significant for soccer audience and are called primary events. Accurate detection of these events can be useful for soccer video summarization. This means that if an observer only reviews this series of events, the result of the match will be recognized and one can enjoy the interesting scenes of the match. Secondary events are those remaining events which are not as significant as primary events for soccer match analysis and cannot affect the match result independently; however a sequence of these events occurring after each other may result in occurrence of a primary and highlight event. Events such as corner kick, foul, player substitution, passing and shooting can be recognized and detected and are useful for technical analysis of players, tactical analysis of teams and their behavior and are used for match statistics provision.

### 2.1 Goal detection

One of the most impressive and interesting events of a soccer match is the goal event. Its significance is such that news broadcasting channels often display only the match scored goals as the soccer summary for their viewers. Moreover the goal detection idea has been paid so much attention in the soccer analysis research.

Goal event detection by dividing video into important elements such as shots and scenes is a popular approach. In many cases, the goal event is detected when a break is recognized in the match, some signs of players cheering are observed or some replays are displayed from different angles captured by the cameras (D'Orazio et al. 2009a). Video shots can contain an event which implies a special concept for the viewer. In Kolekar et al. (2009a), the semantic concepts are extracted using a hierarchical tree which classifies all interesting shots of soccer match in a top-down manner based on a sequential association concept. The interesting clips which are labeled by semantic concepts are located in the lowest level of the tree as detected events. In each level of hierarchical tree, concepts such as replay or non-replay, field-view or non-field-view, field-view classification based on location (long, straight or corner) and player close-up classification distinguishable with the crowd close-up are extracted and players of team A and B are recognized. Finally, based on the results obtained from different semantic levels and by mining their semantic concepts considering event sequences which are dependent semantically, the rules which are stored in sequential association rule-base are applied and the interesting events as well as goal event are extracted.

Visual and audio features of the video are the main clues for extracting semantic information and are vastly used in the literature for this purpose. In Kim et al. (2005), Otsuka et al. (2005), Cheng and Hsu (2006) the audio features are extracted from the video stream considering the increase of audio energy in highlight events ignoring the high cost of visual features extraction. In the next step the important events are selected among the detected events by keeping the positive candidates and ignoring the negative ones. At last the goal event is detected using acquired information from feature extraction methods using statistical models such as hidden Markov model. In Ping and Qing (2009), Leonardi et al. (2004) the audio and visual features are extracted simultaneously. The audio cues including applause of the spectators and sport commentator excitement, and the visual cues are detected and used for shot classification. After modeling the different levels of extracted features and the regarding semantic concept mining, and considering the event sequences which are related to each other semantically such as ball entering the goal and spectators' cheering, the rules available in sequential association rule-base are applied and the goal event is detected (Kolekar et al. 2009a; Shyu et al. 2008). Finite State Machines can also be applied for modeling the events based on specific rules with high precision. In Poppe et al. (2010), this method is used while detecting color, audio and motion features and classifying the shots by SVM and Radial Basis Function Kernel. The method proposed in Xie1 et al. (2007), Wang et al. (2004) is also automatic and can reduce the negative (non-event) candidates using rule-based and distance-based methods and enhance the identification of positive candidates (events) without pruning. The proposed algorithm is in such a way that first some audio, visual and temporal features are extracted and shots are detected by parsing the video. In the next step distance-based mining is performed and the problem of large number of negative candidates compared with positive candidates is discussed and solved. Afterwards rule-based mining is applied and all the candidates are divided correctly into positive and negative classes using C4.5 decision tree. Finally tree structure is created in the test phase and the goal event is detected.

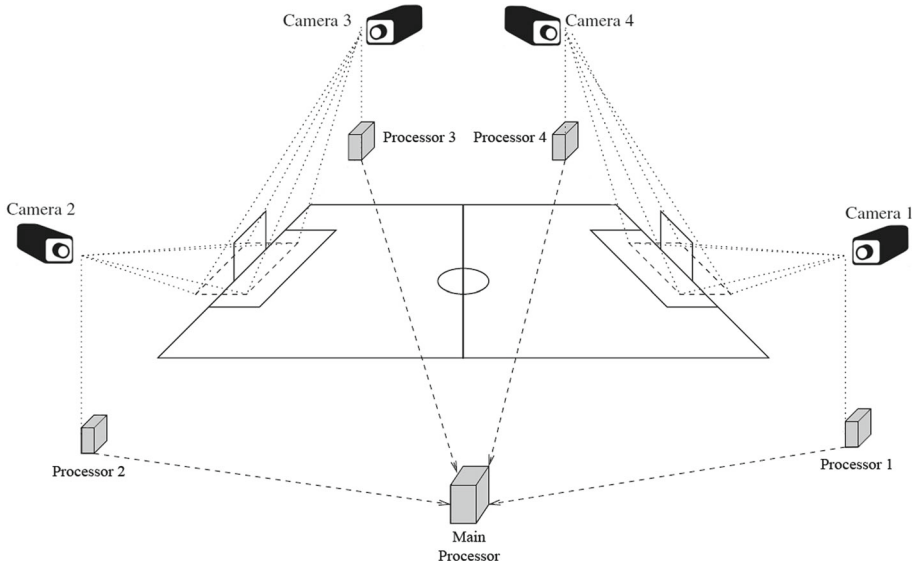


**Fig. 3** A schema of 4-camera system deployment for goal detection

In addition some systems provide training data for their algorithm and train it for some iteration for event detection. A framework is suggested in Wickramaratna et al. (2005) for goal event detection in soccer video based on neural networks. This framework suggest a solution for goal event detection in which multi-objective analysis has been combined with the feed-forward neural network abilities for decreasing the generalization error and the necessary detection has been performed. Likewise the work in Gao et al. (2009) has applied SVM classifier for attack detection. HMM modeling as a popular approach was used in Qian et al. (2011), Kang et al. (2004). After shot classification using SVM, temporal transitions of mid-level semantics and the overall features of an event clip were fused using HMM to determine the type of event. This method was compared with DBN (Dynamic Bayesian Networks) and CRFM (Conditional Random Field Model) as similar statistical models and outperforms these methods for goal, shoot, foul and placed kick by 9.85 and 11.12%, respectively.

Usage of multiple cameras is popular for realtime detection of highlight events in the literature. Besides, some doubtful events often occur during a soccer match than cannot be easily judged by the referee. An automatic and realtime system will be appreciated by the referring committee and be used as a decision support tool by the referees during a soccer match. In D'Orazio et al. (2009a) a realtime system is suggested for goal detection. Four cameras with high frame rate are installed in two sides of the goals according to Fig. 3. Four processors analyze the video captured by the cameras and detect the ball position in realtime mode. The result of analysis is sent to a central supervisor which alerts the goal event to the referee in company with its probability. Two cameras register all events in each goal and the goal event is detected by the local processor, then the result will be provided for the supervisor processor.

Eventually, one of the current used approaches for event detection is statistical models implementation. The method which is used for semantic analysis in Huang et al. (2006) is dynamic Bayesian network. This method is applied for graphical modeling of conditional dependencies between random variables. Nodes are assigned to each of the low-level and

high-level features of the video. The mid-level features are also considered for occurrence of partial scenes which are used for high-level features detection. Afterwards the temporal dependencies between random variables of the nodes are calculated and then cause and effect relations between the nodes are determined. In this way the Bayesian network model is extracted. The mentioned approach is used in Huang et al. (2006) for goal, booking, penalty and corner detection.

### 2.2 Detection of other primary events

Beside the goal event, other significant events such as penalty, booking and offside are considerable and play a major role in soccer video analysis and video summarization.

- **Offside**
  In the research done in D'Orazio et al. (2009b) realtime offside detection is discussed for a soccer match. The criticality of offside event is due to the fact that there are some doubtful situations in which the referee rejects a scored goal and announces an offside kick. In general, offside detection is complicated because of simultaneous monitoring of multiple issues such as ball and players location. In the work of D'Orazio et al. (2009b) six fixed cameras are installed in two sides of play field (three cameras each side) in order to ignore the perspective errors and occlusion caused by players overlapping each other. Afterwards the player who have passed the ball, the location of players exactly at the time of ball passing, and the location of the player who receives the ball are determined and the offside rules are applied for the correct offside detection. Another approach used in Hashimoto and Ozawa (2006) is about team formation analysis by classifying uniforms. Offside line is calculated by projecting the world coordinates of an offside line on an input image and 3D coordinates and the trajectories of the ball are calculated in world coordinates from the plane coordinates of a ball in multi cameras, all for making an offside judgment.
- **Penalty**
  For penalty event detection, the authors of Huang et al. (2006) have modeled the cause and effect relations with dynamic Bayesian network and proposed a solution for penalty event. Mid-level features such as close-up of players, spectators view and replay scenes were used for penalty event detection. Meanwhile the authors of Yong and Tingting (2009) have integrated multiple feature fusions and applied a SVM classifier to detect penalty and goal events. Audio features are exploited in this research by applying a speech recognition engine.
- **Booking**
  Dynamic Bayesian network was used in the same way for booking event detection. The scene of receiving yellow card and especially red card is among highlight events of soccer match and is used for match summarization. Based on the solution suggested in Huang et al. (2006) we can recognize the booking event by viewing close-up of the referee and one or more replay scenes. For more precise detection we can detect the yellow or red card object in the scene. In this respect, card event is detected by aligning textual events with the real match events in Xu et al. (2008a,b), Hu (2010). On the other hand, probabilistic Bayesian belief network (BBN) method is applied in Kolekar (2010) for automatic indexing of excitement clips, classifying the events through a hierarchical tree and detecting the card event.

### 2.3 Secondary events detection

As said before, the remaining events which possess the second place of importance are categorized as secondary events. In many cases, these events form the start of primary events. Researches done in this respect are so limited because they are mainly focused on the primary events.

The system proposed in Shyu et al. (2008) first extracts the audio and visual features and then analyses each extracted shot of video in terms of important events and concepts. The corner event is detected semantically based on the related shots. Dynamic Bayesian network was used for corner detection in Huang et al. (2006) as mentioned before and low level semantic cues such as parallel lines, texture density and short movement of camera are extracted. Afterwards the high level cues resulting from the previous step are extracted which include goal post, crowd behind the goal and camera panning. Also in Yu et al. (2009) the shot boundaries for foul and shoot events where extracted based on semantic concepts extraction from game log and using mid-level and object-level features.

### 2.4 Systems in support of event detection

In some works a solution for classification of extracted features and feature reduction was proposed for optimization of feature extraction problem and facilitating the events detection. MMP[1] method was used in Shen et al. (2008) which is a linear transform for dimension reduction and is based on two principles. Firstly the samples of the same class in the transformed space shall be similar to the maximum possible extent; secondly the samples of different classes in the transformed space shall be different to the maximum possible extent. Combination of these two optimization problems results in solving a package which is in fact a dimension reduction with special value method, however these special values cannot be extracted directly from the feature vectors but can be calculated by solving the said optimization problems. In another hand, the calculation method of special values for dimension reduction is an independent method from the samples class, but in MMP method the dimension reduction is performed according to the samples class. For solving non-linear problems, KMMP method is proposed which first applies a kernel function and maps the features from primary space to the secondary space. In the next step MMP is applied on the features in secondary space. In fact the proposed solution is independent from domain and is applicable in general for optimized extraction of features and rapid detection of events.

The works done in the field of highlight event detection are summarized in Table 1.

As shown in Table 1 and according to the methods discussed in this section, several methodologies have been used for detecting the important events especially the goal event. The methods fusing multiple features specially text sources appear to detect the events with precision near 100%, while hierarchical approaches based on shot classification also appear to be so efficient and their results are between 85 and 90% precision. Instance based approaches which are depending on training data such as Bayesian Networks and Neural Networks generate results with about 80% of preciseness. One interesting result was related to the work in Kim et al. (2005) which detected the goal event only using audio features with 87% precision while a six-camera system could detect the offside event with less accuracy of 82%. On the other hand, result of data mining-based approaches varies between 70 and 85%, depending on the type and number of rules applied in the concept modeling.

[1] Modality Mixture Projections.

**Table 1** The review of soccer highlight event detection works

| Reference (year) | Data sources | Methodology | Detected events | Image stream |
|---|---|---|---|---|
| Qian et al. (2011) | Visual, text | HMM<br>Dynamic Bayesian network<br>CRFM | Goal, shoot, foul, placed kick | Broadcast image |
| Poppe et al. (2010) | Audio, visual | Finite state machine<br>SVM | Goal | Broadcast image |
| Yong and Tingting (2009) | Audio, visual, text | Weighted SVM | Goal, penalty,corner | Broadcast image |
| Ping and Qing (2009) | Audio, visual | Multi-clues detection rules | Goal | Broadcast image |
| Yu et al. (2009) | Audio, visual, text | Instant semantics acquisition(ISA) | Goal, offside, foul, shot | Broadcast image |
| D'Orazio et al. (2009a) | Visual | Background subtraction<br>Circle Hough transform<br>Discrete wavelet transform | Goal | four fixed cameras |
| Shyu et al. (2008) | Audio, visual, text | Subspace-based multimedia data mining | Goal, corner | Broadcast image |
| D'Orazio et al. (2009b) | Visual | Unsupervised clustering of players trajectory analysis | Offside | Six fixed cameras |
| Xie1 et al. (2007) | Audio, visual, temporal | Representative subspace Projection modeling,<br>C4.5 Decision tree model. | Goal | Broadcast image |
| Huang et al. (2006) | Audio, visual | Dynamic Bayesian network | Goal, corner, booking, penalty | Broadcast image |
| Kim et al. (2005) | Audio | Hidden Markov model | Goal | Broadcast image |
| Wickramaratna et al. (2005) | Audio, visual | Neural network framework | Goal | Broadcast image |

## 3 Semantic detection for video summarization and retrieval

In recent years, the digital data of sport videos are spreading rapidly and automatic acquisition of important segments is an inevitable issue (Nitta et al. 2009). This issue should be addressed in aspects of efficient video summarization and video retrieval.

### 3.1 Video summarization

Nowadays, summarization systems have gained special attention because of saving time for users and decrease in the amount of data transfer. The main goal of summarization is to substitute a long video stream with few numbers of selected scenes by ignoring some unimportant video frames and decreasing its duration so that the main content is maintained and the remaining frame sequence keeps the existing semantic relations (D'Orazio and Leo 2010). The output of summarization systems can be used in different applications including TV broadcasting, website and mobile services for the users who do not have sufficient time for viewing the whole match. The recent proposed summarization systems generally work in offline mode and the match summary is provided with some delay after ending of the match. Therefore the new research efforts are focused on creating the match summary automatically and instantly with more accuracy with respect to users preferences and taste based on different formats and qualities.

In case of soccer videos, we seek for extracting interesting events of the whole game such as goals, bookings, shots on target, penalty, etc. Automatic detection of highlight events and semantic interpretation of the scenes is a challenging task in soccer video summarization. Fortunately this can be done by extracting features in different semantic levels (Ekin et al. 2003). For instance, we can perform event detection using replay scene, close-up of the referee, audience cheering and increase in commentator's speech energy.

Basically the video summaries consist of some suitable audiovisual features for representing a useful abstract from the video stream content to the user. However there is no standard available for the type of these features to be used in the summary. In Money and Agius (2007) four types of audiovisual cues are introduced to be used for presentation to the user. These cues are key frames, video segments, graphic cues and textual ones. In another hand, according to the research in Nitta et al. (2009) sport video summary types are divided into dynamic and static categories. In dynamic summaries, three approaches of selecting general highlight scenes as basic criterion, selecting local highlight scenes as Greedy criterion and play-cut criterion are used for detecting important scenes and finally they are sorted in time order. For static summaries an effective presentation of metadata available in MPEG7 is created in tree structure from sport match so that the user can select the target scenes from a list of key frames. Therefore the static approach is 2D form in contrast to the dynamic mode and only the key frames of important events are arranged in order.

A combination of audiovisual features is used in Duan et al. (2003) for summarization in which rule-based methods are applied on the extracted information during event detection. The events of goal, foul and shot on goal for soccer and tackle, goal, mark and behind for Australian football were extracted with an acceptable precision (Tjondronegoro and Phoebe Chen 2009). The main benefit of the proposed method is "knowledge discount" which means optimized usage of a low-level knowledge in multiple domains so that the method remains independent to any special domain. As a result the proposed algorithm works robustly for all sports that have play-break structure and their audiovisual specifications are the same. From another point of view, a collection of labeled highlight clips can be applied in video summary for so many applications such as important events browsing, indexing and retrieval.

Kolekar (2010), Kolekar et al. (2009b) have introduced a method for soccer video semantic analysis and summarization by identification of concepts using Bayesian belief network in which the highlights of the game are detected from various sub-clips using audio features while applying generated rules and domain knowledge. Also in Kolekar et al. (2009a) the summarization of long video streams is performed by detecting highlight events using audio features and labeling semantic concepts by creating a hierarchical tree of different shots. In this method each of the semantic concepts consists of sequential related events.

The last but not the least is textual sources that have been used in literature for soccer video summarization. Yu et al. (2009) suggests a method by which the users who do not have access to the live broadcast video from TV can enjoy the clips related to match highlights with minimum latency via internet or their mobile phones. Eventually they are provided with the video summary according to their preferences based on events such as goal, booking, shoot and offside. In Xu et al. (2008a) summarization is performed using webcast text and ontology based methods without using video content. In this approach the soccer highlights are presented based on analysis and alignment of webcast text with broadcast video. Also in Qian et al. (2010) a novel approach is introduced in which visual features are used along with the text of annotated video to extract necessary semantics and identification of pertinent segments for summarization. Again by using text sources in Hartley and Zisserman (2000) a system for attack event detection using analysis and alignment of long shot images is proposed.

### 3.2 Boundary detection in soccer video streams

Generally video frames are the smallest elements of video which consist of one image. Frames can be used for understanding low-level concepts but cannot be processed for extracting high-level semantics such as match statistics analysis, ball detection and player tracking. Detection of a sequence of frames which are captured by one camera during a scene is necessary for this purpose which is called a video shot. Shot change detection methods are usually focused on usage of features extracted from DC coefficients, motion and macro block information in order to detect the important changes in the scenes. Basically the shot detection algorithms can be designed and used as an infrastructure for event detection mechanisms. Video shots are known as basic elements for indexing and table of content creation for video and establish the real and basic video physical layers.

Soccer video shots are used along with a series of logical rules as primary knowledge for modeling event structures in videos and for event boundary detection. In Wickramaratna et al. (2005) an automatic method is proposed in which the shots are classified into non-hitting, in-field and out-field categories using three-layer feed-forward neural network. Shot boundary detection is performed in Chen et al. (2005) based on pixel-histogram comparison, segment map comparison and object tracking using a series of low-level visual and some low-level audio features in frequency and time domains. Moreover, a top-down method for semantic classification of video shots is proposed in Xu et al. (2009). Low-level features similarity and its mapping to non-parametric features such as Motion Vector Field Model (MVFM) and Color Tracking Model (CTM), in company with some color features such as histogram are used for shot length measurement. Finally the shots are classified by using human knowledge and SVM classifier method which is a machine learning technique.

Beside the application of visual features in boundary detection, webcast text analysis is another known approach in this domain. In Xu et al. (2008a) the boundary of shots is determined by detecting events start and ending time using webcast text and Conditional Random

Field Model (CRFM) method. The shots are divided into in-field, out-field and close-up in this research. Hidden Markov Model (HMM) is used in Xu et al. (2008b) for event boundary detection after detecting the time of events. The precise time of events is labeled by matching the video clock with clock digital characters and then the sample digit pattern is detected automatically. In the same way in Eldib et al. (2009) first the shot boundaries are detected using G distance[2] and H distance[3] considering a special threshold and then shots are classified into general view, middle view, close-up and audience groups. Beside the application of game log information as semantic concepts, the work in Yu et al. (2009) has applied dominant color and motion intensity in support of video clip boundaries detection.

According to the structure and content of sport videos, two main features of "play-break" and "replay" has been vastly used in events boundary detection which will be discussed in this section.

### *3.2.1 Play-break detection*

Play-break detection is a new approach which is recently introduced in semantic analysis domain. It can be used in different applications such as summarization of all sport videos having play-break structure like football and basketball. Play means a video sequence in which the ball is circulating in the normal flow of the game. Break means the period in which the game is stopped and usually the ball lies outside the field. By this view the long sport videos will be divided into smaller sequences. For instance, in soccer game when the referee's whistle is heard it means that a foul has been committed or the ball is outside the field. For detecting this issue, the replay and close-up scenes can be used depending on the length of the break. Sometimes text display, logos and advertisements are also shown in the break period.

Moreover the break ratio is a more robust and flexible criterion in comparison with the break length measurement and eases the process of finding break sections. For solving the problem of inserting artificial text within the video for events description, the play-break feature can be used in which the textual description can be automatically inserted during break sections or at the beginning of play section of the related event (Duan et al. 2003).

A schema of semantic analysis is introduced in Duan et al. (2003) for creating video summary using play-break method. At first the play-break sequences are segmented based on camera view classification and replay detection. A sample of said classification is shown in Fig. 4. Afterwards the mid-level features are extracted and statistics of each play-break such as its ratio is calculated. These statistics are trained for each highlight and then the highlights are classified and the video summary is prepared. Likewise, features such as the amount of grass pixels in a frame and the white lines in the goal area are used in Tjondronegoro et al. (2003) for play-break recognition. In Xie et al. (2003, 2004) HMM method is applied in the same way for play-break detection and the work in Wang et al. (2005b) have used a multimodal multilayer statistical inference framework that recognizes play and break segments using dynamic Bayesian networks.

### *3.2.2 Replay scene detection*

Nowadays broadcasting companies display replay of exciting and important scenes to emphasis on special events with full details. The replay scene mainly consists of a slow motion dis-

[2] Difference between dominant colored pixel ratios of 2 frames.

[3] Difference in color histograms based on HSV color space.

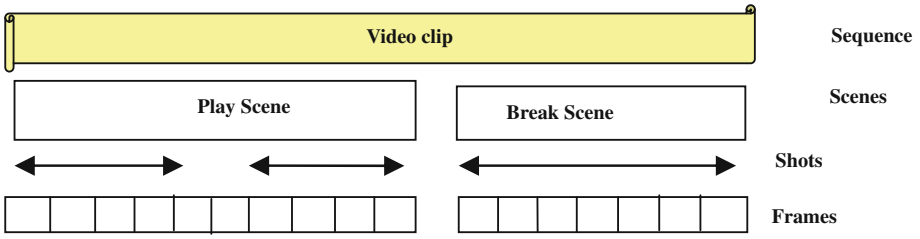


**Fig. 4** Video classification based on play-break

play of an interesting event and sometimes a logo (matches special sign or sponsors trademark lasting for some frames) which is used at the beginning and end of the replay. Some works use replay scene in different fields specially boundary detection and often in event detection and summarization systems because of its explicit visual features and specifications. Two points of views are considerable in the replay scene detection. First one is using the slow motion pattern and the second one is logo detection.

In replay scene detection based on slow motion feature, change of shots and its change frequency is detected and compared with the shot change frequency of normal frames (Pan et al. 2001; Wang et al. 2004). This is performed according to specific thresholds to decide about detection of replay scene. Also kinematic features are used in Yang et al. (2004) to locate goal events since they are generally followed by slow motion replays. In cases that the broadcast videos contain logo, replay detection based on logo is applied (Kolekar et al. 2009a; Eldib et al. 2009). As an instance, summarization is performed by replay and rule-based goal and attack detection in Eldib et al. (2009) similar to the distance based method used by Ariki et al. (2006) for summarization. This detection is possible by boundary detection based on goal mouth, shot classification, replay detection and scoreboard detection. Another sample is the proposed system of Yang et al. (2007) in which the authors used the goal mouth along with audio energy for detecting important events such as goal and attack. Similarly the work in Xu et al. (2008a) used Special Digital Video Effects (SDVE) and an algorithm of video frames comparison for logo detection. Finally the method used in Xu and Yi (2011) detects the logo first by extracting logo template from soccer video and then finding logos by comparing the similarity between candidate images and the template using k-means clustering algorithm. The condition of logos to be paired is applied as well in this method.

### 3.3 Smart display of soccer videos on small LCDs

Advances and improvement of multimedia signal processing has led to vast usage of portable devices with small LCDs. Dissatisfactory of people viewing images on these LCDs has resulted in performing research in regard of determination of user region of interest (ROI) and its magnification. Small LCD images are not sufficiently clear due to the fact that the videos are created for displaying on normal TVs with larger dimensions. Especially small objects such as ball in soccer video are hardly recognizable when displayed from far distance. The main goal in this domain is determination and display of user region of interest in the videos which is tightly related to video content recognition. Region of interest refers to the segment of image which has more importance to the viewer and usually contains the ball due to criticality and dependency of the soccer events to this object (Pei et al. 2009). A sample of region of interest is displayed in Fig. 5.

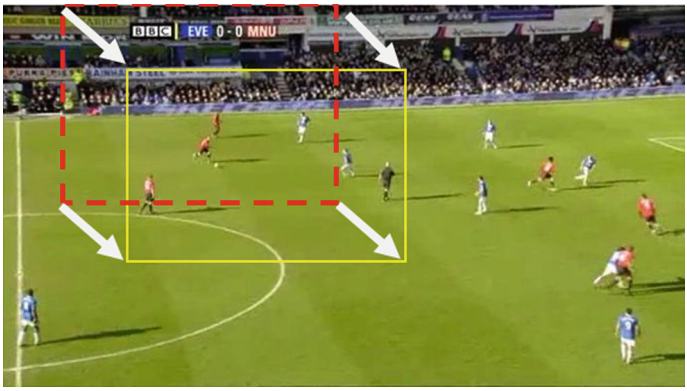


**Fig. 5** Region of interest in soccer video

One of the important challenges in this domain is simultaneous occurrence of multiple events which makes difficult the task of region of interest selection. The next challenge is detection of frames in which no important event occurs and there is no need for magnification.

The method proposed in Seo et al. (2007) displays the ROI by applying suitable scale for the user so that the long shot image is viewed in larger size. This task is done in three levels of field color learning, shot classification and ROI detection. Also in Hu et al. (2010) the ROI is detected based on UIO (User Interested Objects). Dominant color ratio and GMM model were used for shot classification, MBR (Minimal Bounding Rectangle) was applied for ball detection, Viterbi algorithm for ball tracking in successive frames and the player's face detection was settled on AdaBoost based face detector with Haar-like features. Decrease of performance in different lighting conditions and dependency of the algorithm to the video quality are the deficiencies of the method proposed in Seo et al. (2007). In addition the system is unable to prepare the result in online mode, therefore the mobile and PDA users cannot access the online soccer events.

### 3.4 Soccer video retrieval

Due to every day increase in number of stored video clips, searching among the media and selecting desired pieces have become difficult and often impossible. The traditional method for searching among non-textual data (voice, images and videos) is the labeling mechanism, in such a way that one or more label is assigned to each file by human and these labels represent the content of the file. Labels are searched later for the user query and the result is returned. The use of labeling for searching has so many limitations. Firstly the labeling task and data entry by the user is time consuming, expensive and with human mistakes. Secondly the number of labels is usually limited and cannot fully describe the file content. Also the files that have been stored before the start of labeling process are unusable and cannot be searched.

Nowadays for eliminating these restrictions, video retrieval methods are used for searching which is a combination of different sciences such as information retrieval, machine vision, machine learning, human and machine communication, data mining, philosophy, event photography and video capturing techniques. In this approach the user submits an image or small clip to the search engine and system returns all similar images and videos which contain similar frames. This is called content-based retrieval in which the internal content of the

images and videos are considered instead of the created labels. Semantic-based retrieval can be considered as a higher level of retrieval in which the semantic embedded in the image or video will be extracted automatically by the search engine.

The video retrieval systems are challenged by two basic issues. First challenge is sensory gap which refers to the gap between the object existence in the real world and its description in the image. Efficiency of capturing devices and their role in image content identification is discussed for solving this problem. Another challenge is the semantic gap which is related to the incompatibility between extracted concepts by a retrieval system and the related concepts perceived by human mind. In another words, interpretation of semantics from image information is different between the computer system and human which should be considered while designing a retrieval system. Different methods and efficient devices of image capturing should be used for filling the sensory gap. In another hand for solving the problem of semantic gap, suitable methods for similarity measure and matching of extracted features from image or video should be applied so that these methods simulate the functionality of human mind in most efficient way. The problem of semantic video annotation is strictly related to the problem of generic visual categorization, like classification of objects or scenes, rather than that of recognizing a specific class of objects (Ballan et al. 2011). In the same way, soccer video retrieval necessitates indexing and annotation of events but for a set of limited number of events and concepts. These concepts can be detected by semantic identification and event detection from videos. In addition, indexing based on video stream semantics is a novel research field and is effective for soccer video retrieval process.

In this regard, a framework for semantic annotation is proposed in Xu et al. (2008b) in which the sport video is retrieved according to the user preferences using webcast text. A method is suggested in Kolekar (2010) for detection and annotation of semantic concepts such as goal, card, goal keeper save, etc. on broadcast videos in which Bayesian belief network is used for determination of visual events dependencies with high level semantics. Replay scene, field view, goal keeper, players, spectators, referee, gathering of players and a set of labels are used in this research for concepts annotation. The detected events are classified and labeled using low level features and based on a hierarchical tree. In this method the events detected within video clips are considered as instances for Bayesian network and the highlight clips are semantically labeled based on posterior probability. In the similar manner, the work in Ballan et al. (2010) has performed the annotation task based on instance clustering, prototype selection and dynamic cluster updating to be presented to the ontology for annotation of new observations. Web Ontology Lang (OWL) and Semantic Web Rule Language (SWRL) were applied for reasoning over semantic concepts.

Another automatic method is proposed in Kolekar et al. (2009a) for soccer video indexing and labeling according to semantic concepts detection based on detection of hierarchical events. These events are extracted from video using analytical combination of low level features with high level semantic knowledge and each sequence of events are classified into a sequential association concept. Also a priori algorithm is applied in this method for complexity calculation of each highlight clip. The semantic concepts used in this approach include "goal scored by team A or B" and "goal saved by team A or B". Likewise in Assfalg et al. (2003) automatic annotation is performed for highlights in soccer video, suited for both production and posterity logging while the knowledge of the soccer domain is encoded into a set of finite state machines.

The authors of Nguyen and Ly (2010) have presented an automatic independent to language system that recommends the users to query events by keywords in image form which are the agents of clusters of stationary on-screen text boxes using HACL (Hierarchical Agglom-

erative Clustering) algorithm. The stationary blocks are detected between two frames by applying canny operator and connected component analysis.

Table 2 represents the novel works done in the field of video summarization and retrieval. The papers which are focused on summarization are using play-break detection and replay-scene detection approaches. Various features of video, audio and text are used and the result is generated for the users. The summaries are still obtained in offline mode and instant summary preparation is expected to be achieved in the future.

## 4 Field object tracking for tactical analysis and statistics provision

Ball and player are the most significant and vital objects of a soccer match because the match result as well as most of the events are dependent to their existence. Difficulties in tracking objects are mostly due to abrupt object motion, changing appearance patterns of object and the scene, non-rigid object structures, object-to-object and object-to-scene occlusions, and camera motion (Yilmaz et al. 2006). In this respect by using high-tech computer science, movements and interactions between ball and players in the field are identified and analyzed so that it would help for solving the analytical problems which are important to referees and coaches. This issue includes team analysis based on technical and tactical behavior, general match analysis and minimization of the faults committed by the referee.

### 4.1 Ball detection and tracking

Ball is in the center of attention in every ball-based sport like soccer. However the players can identify the ball easily according to its color and appearance, but automatic acquisition of precise location of the ball in the field may be difficult due to rapid movements of the cameras and field partial views. Ball detection and tracking is challenged by various issues from physical specification point of view, because rather small size of the ball comparing to other field parameters also its high speed (especially during attacks) will cause the ball to appear in different sizes even colors and in blur or unclear state (Ren et al. 2008). In addition ball similarity with other field objects such as the region near field lines, player's regions, player's jersey or some circular shapes in the scoreboard, make the ball detection more difficult (Pallavi et al. 2008b). Also the ball color varying in heavy weather and lighting conditions and the occlusion problem (ball merged with player) are the deficiencies of detection and tracking (Choi and Seo 2005).

A method is proposed in Ren et al. (2009) for soccer ball tracking with occlusion occurrence probability in which the occlusion problem is solved using eight fixed cameras according to a model-based methodology and back-tracking. First the ball candidates are determined using color, size and shape features. Each tracking is fulfilled by means of Kalman filter and calculating ball existence probability. This measure is corrected by back-tracking results. Finally a simple 3D tracking is performed in which the ball 3D location is mapped to 2D for more efficient recognition and more powerful tracking. However using multiple fixed cameras around the field will provide a wider view, but by increasing the number of cameras, heavier CPU process and more equipment will be necessary in comparison to the usage of broadcast videos.

One of the famous approaches for ball detection is the background subtraction method. A system is proposed in Joo and Chellappa (2007) in which first the region of moving objects in terms of blobs are detected for ball visual tracking using background subtraction, afterwards the relations between objects are determined by a binary mapping in the object region, also a

**Table 2** The review of video summarization and retrieval works

| Reference (year) | Data sources | Methodology | Detected events | Application view |
|---|---|---|---|---|
| Ballan et al. (2010) | Visual | Web ontology language (OWL) | Shot-on-goal, placed-kick, foul | Video semantic annotation |
| | | Semantic web rule language (SWRL) | | |
| | | Visual instance clustering | | |
| Tjondronegoro and Phoebe Chen (2009) | Audio, visual | Logical rule-based models | Goals, shot on goal, foul | Play-break detection, video summarization |
| Kolekar et al. (2009a) | Audio, visual | Sequential association mining | Goals, saves, corner | Replay detection, |
| | | | | Video indexing and summarization |
| Kolekar et al. (2009b) | Audio, visual | Probabilistic Bayesian belief network | Goals, saves, booking | Excitement clips detection |
| Nitta et al. (2009) | Audio, visual, text | Tree structures on rank of play scene | Using MPEG7 events | Video abstraction |
| Eldib et al. (2009) | Visual | Dominant color, histogram intersection methods | Goal, attack, foul, booking | Replay detection, |
| | | Rule-based classifier | | Video summarization |
| Gao et al. (2009) | Audio, visual | Hierarchical agglomerative clustering | – | Video summarization (for general domain) |
| | | Two-level redundancy detection | | |
| Xu et al. (2008a) | Visual, text | Probabilistic latent semantic analysis (PLSA), | Goal, corner, shot, card, free kick, offside, foul, substitution | Video summarization |
| | | Conditional random field model (CRFM) | | |
| Xu et al. (2008b) | Audio, visual, text | Statistical model, | Goal, corner, card, free kick, offside, foul, shot, save, substitution | Video semantic annotation, retrieval, |
| | | Hidden Markov model (HMM) | | Personalized text and video summarization |
| Ekin et al. (2003) | Visual | Cinematic features, dominant color | Goal | Video summarization |

simple Kalman filter is used for estimation. Whereas divide and merge is possible for visual tracking of the objects, the problem of data relations as the problem of minimal cover for edges weighting is discussed. Similarly in Ren et al. (2010) ball is detected in a two-tier architecture in which single-view and multi-view processing are implemented. After background subtraction, for ball filtering process, velocity and longevity features, along with size and colour are employed to discriminate the ball from other objects. In the next step, ball 3D position is estimated using multi-view of eight cameras.

Some obstacles prevent the 3D location estimation of moving objects in broadcast videos, such as ball floating in the air, undetermined intrinsic and extrinsic parameters of the cameras and difficulty of 3D information extraction from just one view. In addition, ball detection in one frame is not always possible due to its small size and fast speed and requires manual human assistance. A method is proposed in Liu et al. (2006) for ball 3D location estimation using broadcast videos for ball tracking in match field in which ball candidates are selected in each frame and then the ball location is determined in sequential frames by Viterbi decoding algorithm. Afterwards, all 3D information estimation is presented without referring to another object with known height in this algorithm. Another system exploits a field model extraction algorithm to extract calibration information in goal and non-goal scenes. The two stage camera calibration approach includes camera position calculation using DLT (Direct Linear Transformation) method and solving projection equations by the non-linear optimal Levenberg Marquardt (LM) algorithm where camera parameters were calculated (Gao et al. 2011). Camera parameters are useful low level features which are good for scene analysis and object matching.

Hough circular transform with neural network classifier is used in D'Orazio et al. (2002, 2004) for ball detection based on the video recorded by the authors' own camera. The experiment is done on the middle shots in this work. Therefore the algorithm fails with high probability if objects similar to ball exist in long shots. The work in Tong et al. (2004) uses a condensation-based method for ball detection and tracking in a way that ball is denoted in coarse to fine process by combination of color and shape features. However this method fails in crowded and complex backgrounds. Kalman filter is used in Yu et al. (2003, 2006) based on trajectory search and evaluation for ball tracking. In this method a set of ball candidates are generated by pre-processing of the video. Color feature is used as the only main information source for ball detection in Yu et al. (2002, 2003, 2006). Therefore this method does not work properly in lighting variations, while Pallavi et al. (2008b) has used ball shape information in addition to the color feature. Morphological methods are also used for ball detection (Hossein-Khani et al. 2011). Image segmentation is done in this approach for play field, field line and ball detection and then the performance of the system is compared with CHT and appears to be about 20% more efficient. The limitation of CHT method is the fact that it is necessary to set the radius of circular shapes according to the ball appearing with different radii in various shot types.

There is a method proposed in Miura et al. (2009) for ball route estimation in a 2D coordination of broadcast TV images. First a graph is created which represents the possible transitions of the ball between the overlapping objects based on their spatiotemporal relations. Then the number of ball route candidates is measured by the graph and the best one is selected by searching for sign of ball existence near each route candidate. One of the advantages of this method in comparison to the others is increase in tracking speed because there is no need to scan the whole image during search for the ball. A disadvantage which can be mentioned is its dependency to the images captured by the central camera and small usage of other scenes which result in failure of the method for long sequences.

There is an algorithm suggested by Pei et al. (2009) for ball routing which is based on extension strategy for ball detection and tracking. In this method in addition to ball detection, the balls occluded by other objects in the frame are also recognized. Four general steps are mentioned for ball detection and tracking. First, ball size estimation using a pre-processing level, then finding the ball candidates and detection of main ball among the ball candidates based on the ball route information and finally route processing. The advantage of this method is about the applying of a decreased set of ball candidates instead of direct ball selection in a frame which allows the occluded and mixed balls to be considered as ball candidates.

## 4.2 Player detection and tracking

Player detection, recognition and tracking are known as important issue in the field of soccer video analysis. In the machine vision domain, tracking moving objects in a sequence of images is an interesting but difficult issue while player movement and objects inside the field and their interaction carry a special semantic and concept. Single and multiple players tracking are useful for detection of many primary events such as goal and offside. In addition performance of a player can be evaluated by considering the movement path, the traversed distance and speed. Also performance of a team and the used tactics during the match can be achieved by player tracking as a function of time.

According to Beetz et al. (2006) player and ball detection issues using broadcast videos can be categorized in three domains: relevancy of color regions, camera parameters decoding (pan, tilt angle and zoom) and players positioning and identification along with ball detection. In this respect, the work in Gedikli et al. (2007) has used special template for calculation of player locations likelihood maps according to color distributions and segmented regions specifications. Also player detection and labeling is proposed based on unsupervised learning in Liu et al. (2007) using broadcast videos. Meanwhile, many approaches have been introduced based on exploiting fixed cameras such as graph-based trajectory (Nillius et al. 2006; Sullivan and Carlsson 2006) self- evaluated weighted graph in both backward and forward direction (Misu et al. 2004) and unsupervised clustering of player uniform histograms (Spagnolo et al. 2007).

As said before, occlusion problem is known as an obstacle for player tracking. It means that while following trajectory of one or more players, an object may prevent the continuous tracking and therefore we need to search for our object in the next frames. A solution to this problem is the usage of multiple camera views for object tracking. While facing occlusion with one camera view especially for broadcast videos, we can compensate the missing views by using more cameras and track the object's trajectory.

Works in the literature are in two categories in this respect. In the first category the images of multiple cameras are used for player tracking and ignoring occlusions. In the second one, due to limitation of data sources to broadcast videos, we should use optimized algorithms and often human assistance for solving the said problem.

### *4.2.1 Using multiple cameras for tracking*

Installation and calibration of multiple cameras in different angles of match field while their intrinsic and extrinsic parameters are determined is a suitable approach for effective tracking of players. Using multiple integrated cameras enables us to succeed studying the difficult and imprecise situations and remove the ambiguities related to cameras calibration.

The work in Du and Piater (2007) has proposed a multi camera player tracking based on particle filters and axis of cameras. In this novel solution, the desired object is not tracked by

one camera, but is tracked in the whole match field by particle filters separately. In general conditions of multi camera systems, if the interaction between the cameras is not considered, object matching cannot be fulfilled, while in Du and Piater (2007) thanks to the feedback reception from tracking of field area, the tracking result will remain consistent in each camera even if they track a wrong object.

3D location estimation of players and ball by means of the central camera which has a wide suitable view is a useful solution for tracking in the match field. In Liu et al. (2006), Misu et al. (2007) the players are found by match field detection and are tracked by particle filter and based on support vector regression. In this research, a novel algorithm for 3D information estimation is presented and automatic ball and player detection is performed to decrease manual labeling operations.

Using the background subtraction method, the system proposed in Poppe et al. (2010) detects the players by applying code-book algorithm. The objects are then classified based on color histogram and Bhattacharyya distance. Later, the information of the objects in different camera view-points is combined by projection on a synthetic top-view playing field. Consequently, the different projections are merged to obtain the trajectories of the players with at least 88% precision.

It goes without saying that player behavior and movement in the field as a function of time is a useful information which assists the task of player performance analysis and improvement. The works in Figueroa et al. (2006), Xu et al. (2005), Du et al. (2006) discussed the issue of player tracking during a soccer match using multiple cameras. The main goal of research in Figueroa et al. (2006) is to determine location of players in each moment of time. A model of players and several morphological operations are applied to solve occlusion and crowd problem. Tracking is done through a graph in which the nodes are adjusted according to the detected blobs from image segmentation. The weight of graph edges is determined using the blobs information. The player trajectory in the image stream represents the distance between the nodes. This task is done based on minimum four fixed cameras covering the whole match field together.

#### *4.2.2 Using broadcast videos for tracking*

However using multiple camera images increases the performance of analysis systems, but this solution is not always available and incurs more cost for soccer video analysis. Another available source is the broadcast TV videos. The broadcastvideo is an edited video where the broadcast sequencefeed is selected from frequent switches among multiplecameras according to broadcast director's instruction. Despite the fact that these images often do not have sufficient quality, have limited view angle and are not suitable for solving problems such as occlusion, this source can be always archived, maintained and will be accessible. By using BSV, the system in Sun and Liu (2009) have detected and classified the players of the teams using Gray Value Top-Hat Transform in field region. Team player classification based on jersey color is another popular approach which has been performed in Vandenbroucke (2003), Xu and Shi (2005) by applying initial training manually.

Considering the importance of goal event in soccer matches, a system is suggested by Khatoonabadi and Rahmati (2009) for automatic tracking of player movements in goal scenes. This system consists of four phases. In the first phase, an automatic algorithm is used for field grass detection which is tested for various types of soccer video images and different conditions. In the next phase, match field lines are identified for player location detection by Hough transform. Whereas using Hough transform is expensive for every single frame, the position of the field lines is estimated by Kalman filter in the other frames. A sample


**Fig. 6** Soccer field lines and regions

of soccer field along with its lines and regions is displayed in Fig. 6. In the third phase for player location estimation in the current frame, its position in the last frame is used in company with perspective transform. Finally the region based detection algorithm is applied for player tracking. In case of occlusion occurrence, histogram back-projection algorithm or a combination of merge-split and template matching is used according to occlusion type around the estimated positions. By mapping the player positions from image plane to play field model, even rapid movements of the players will cause little difference in the play field model space and player position estimation will be performed more precisely. Players are tracked by searching in the output of region-based detection algorithm.

Graph-based approaches play a major role in player trajectory detection and occlusion removal (Pallavi et al. 2008a; Miura and Kubo 2008). According to this approach the research in Pallavi et al. (2008a) first classifies the shots in three categories of close, medium and long and then tries to detect the players in long shots. The player candidate regions in long shots are detected by filtering non-player pixels. Afterwards the weight graph is created so as each node represents a player region and each edge shows a link between players in each frame to the players in the next two consecutive frames. In another word, first player position is determined in each frame by removing non-player regions. The remaining pixels are classified and then the region-growing algorithm will identify the player candidates. The weighted directed graph is created when player candidates are available according to a graph node in a way that each graph edge is connected to its candidates in a frame and its candidates in next two consecutive frames. Finally dynamic programming is applied to search for each player trajectory and tracking is performed. The proposed method is robust against different lighting and field conditions and achieved average of 96.5% recall and 90.84% precision.

In some papers the player tracking issue is discussed within a video semantic analysis framework. A multi-layer framework is presented in Xu et al. (2009) for sport video analysis using middle level features (visual, audio) which is able to solve the problem of semantic gap between low level and high level features. SVM classifier for shot classification in conjunction with field segmentation is used for primary detection of the players and start of tracking. Afterwards, particle filter solution is improved by integration of Support Vector Regression (SVR) with Mont Carlo method. The suggested SVR particle filter is applied for multi-player tracking in sport broadcast videos.

Using learning algorithms and training them in some iteration for obtaining optimized results is another popular approach for solving player tracking problem. Method of detec-

tion, labeling and tracking of players is described in Liu et al. (2009b). In this research, player detection is performed using dominant color method based on background subtraction and Haar features according to Adaboost classifier and solved the problems to some extent. The whole process of algorithm is fulfilled in two phases by twice processing of video images. These phases are called training and test. In the training phase, first the video images are processed and the images are classified into four views by extracting color histogram, dominant color and using a decision tree. In the test phase for player detection in match field, first the field region is separated from the general view and then the image is processed with Adaboost classifier with different scales. Usually the overlapping regions of players are detected multiple times and a rectangle is drawn around each player. Labeling of players is performed by learning player appearance in unsupervised manner.

### 4.3 Tactical analysis

Regarding the fact that nowadays the soccer issues are paid so much attention and the knowledge related to this sport field has initiated a new rout in science, therefore soccer analysis and efforts for designing automatic analysis systems have been extended and is in center of attention for coaches, analysts and soccer clubs. In general there are three kinds of analysis models for a soccer match: player performance analysis, team performance analysis and whole match analysis.

In player performance analysis, their behavior and movements are tracked and discussed. For instance, the number of goals scored, received cards, committed fouls, ball possession, number of passes, covered distance, etc. for each player has to be detected. This kind of analysis is for purpose of comparing player capabilities and their progress during the soccer matches. Team tactic analysis is about considering issues such as team formation and player positioning in the field, attack and defense strategies, and team strengths or weaknesses. In this respect, several useful information are extracted based on the relationships between trajectories of 22 players and a ball and the performance of several players is evaluated by considering the interactions between them (Kang et al. 2006). Finally in whole match analysis, the statistics related to highlight events such as goals, cards, attacks, ball possession, etc. are extracted which is paid so much attention by soccer fans and broadcast TV channels.

The soccer tactical analysis is composed of considering player personal behavior and interaction between players and the ball, also locating the field region in which an event occurs, while the soccer coaches cannot perceive all technical and tactical specifications during a match. In addition during the action of dribble, the player and ball are close to each other so distance and trajectory parameters are small because they both cover a short distance. In contrast during passes, the distance between ball and player can be different depending on the type of pass (short, long).

In Zhu et al. (2007, 2008) the attack event is divided into four sub-events of goal, shoot, corner and free kick and the attack tactic is divided into direct and dribbling for individual mode and to unhindered and interceptive attack for cooperative mode. Attack event extraction is performed by using long shots and webcast text based on broadcast videos. Afterwards the information from player's trajectory is combined with information from play active regions using the detected lines and competition network, so the tactical patterns are extracted. The long shots provide a wide view for detecting whole attack scenario and enhance multi-object tracking. The tactical patterns based on attack events are summarized into two individual groups. The first group is based on route pattern recognition which is related to the attack rout in the soccer field such as side and center attacks for which the work in Masui et al. (2010) have proposed a solution independent of tracking objects. In this method, visual pat-

terns and player route patterns are detected for recognition of the area of players' distribution with more than 86% of precision. The second group is about interaction pattern recognition and is related to interaction between ball and players such as dribbling attack. Also in Taki et al. (1996) a soccer team tactic analysis is conducted based on the notion of minimum moving time pattern and dominant region of a player. The positions of players were estimated from the soccer game image sequence which was captured by a multiple and fixed cameras system, and then transformed to real soccer field space using camera calibration technique. Also in Zhu et al. (2009) a system for attack event detection using analysis and alignment of long shot images with webcast text is proposed.

Match statistic provision as another product of match analysis is performed in Yu et al. (2009). An automatic analysis system is proposed for supporting coaches using a set of heuristic flexible rules. This research is focused on extracting statistics related to goal, pass (correct, incorrect), shoot (on target, off target) and offside in a soccer match. Additionally based on players and ball coordinates and analysis of sequential frames, a large number of personal and cooperative events are detected which reveals the strength and weaknesses of players.

Likewise the method proposed in Jiang et al. (2007) extracts information about the strategies of defenses and attacks from a soccer video which is done by scene classification and analysis. Video analysis is performed in three levels of information extraction: object level including object tracking, scene level including scene classification and events level including event detection and their summarization. First the match field region is identified in each frame. In each of the field views, features extracted from view classification, middle line identification and overall motion are considered for scene semantic information extraction. This data is used as middle-level information and finally the statistical results are obtained based on position extraction for the users.

Considering the complexity of team and players' analysis, the research in this field is going to be extended in the future. Few dimensions of this field are surveyed and most of the works have performed a combination of team and player technique and tactic analysis or match statistics extraction. Therefore none of them are certainly focused on extracting detailed information for the use of soccer fans and coaches and so many tasks are still performed manually. The solutions related to important issues such as player possession duration, advanced match statistics presentation and instant acquisition of defense and attack strategies should be improved significantly and the need of human intervention should be decreased. The main works in the field of ball and player detection and tracking are presented in Table 3. Also we summarize the papers that work on tactical analysis and statistical provision in this table.

## 5 Multimodal features and different sources for semantic extraction

Generally the soccer video analysis approaches are classified in two groups based on their using data source. First one is analysis based on video content and another one is based on an external source. Meanwhile the fusion approaches are also available in the literature. Most of the current approaches identify specific events from the audio, visual and textual features directly by image, text and audio processing based on the video stream content. If the analysis is made just by using the video content, the low-level features are extracted and different models are presented for event detection, semantic extraction and in general achieving the high-level features. Due to the difficulty of video analysis based on features extraction, events extracted based on external sources can be an alternative solution. Limitation in number of

**Table 3** The review of works in the field of object tracking for tactical analysis and statistics provision

| Reference (year) | Data sources | Methodology | Image stream | Application view |
|---|---|---|---|---|
| Gao et al. (2011) | Visual | DLT (direct linear transformation) | Main camera | Prior knowledge for object matching |
| | | FPC camera calibration | | Camera calibration |
| | | Levenberg Marquardt (LM) | | |
| Hossein-Khani et al. (2011) | Visual | Morphological operation | Broadcast image | Ball detection and tracking |
| | | Kalman tracker | | Corner event condition checking |
| Ren et al. (2010) | Visual | Gaussian mixture model | Eight cameras | Player detection, |
| | | On-line K-means approximation | | |
| | | Histogram-intersection method | | |
| | | Kalman tracker | | 3-D ball positions through occlusion |
| Poppe et al. (2010) | Visual | Color histogram with Bhattacharyya distance | Six static cameras | Ball and player detection |
| | | Canny edge detection | | |
| | | Hough transform | | |
| Masui et al. (2010) | Visual | Background subtraction | Broadcast image | Tactic analysis |
| | | Rout pattern extraction | | |
| | | KNN algorithm | | Attack type detection |
| | | K-means algorithm | | |
| Abreu et al. (2010) | Visual, text | Heuristics rules | Broadcast image | Tactic analysis |
| | | Sequential time frame analysis based in cartesian coordinates | | Match statistics |
| Sun and Liu (2009) | Visual | Gray value top-hat transform | Broadcast image | Player detection and player type recognition |
| | | Template matching method | | |
| | | Morphological operation | | |
| Liu et al. (2009b) | Visual | Dominant color based | Broadcast image | Player detection, labeling and tracking |
| | | Boosting detector with haar features | | Solving occlusion problem |
| | | Markov chain Monte Carlo (MCMC) | | |

**Table 3** Continued

| Reference (year) | Data sources | Methodology | Image stream | Application view |
|---|---|---|---|---|
| Khatoonabadi and Rahmati (2009) | Visual | Hough transform<br>Kalman filter<br>Region-based detection algorithm<br>Histogram back-projection<br>Merge-split method<br>Template matching method | Broadcast image | Tracking players ,<br>Solving occlusion problem |
| Zhu et al. (2009) | Visual, text | Trajectory analysis<br>Multi-object trajectories<br>Weighted graph representation<br>Temporal–spatial interaction analysis | Broadcast image | Tactic analysis |
| Ren et al. (2009) | Visual | 3D trajectory modeling<br>Kalman filter<br>Motion analysis | Eight fixed cameras | Ball tracking |
| Miura et al. (2009) | Visual, central camera | Transition graph<br>Temporal–spatial interaction analysis | Broadcast image | Ball tracking in occlusion position |
| Xu et al. (2009) | Audio, visual | Color tracking model (CTM)<br>Support vector machine (SVM)<br>Motion vector fields (MVF)<br>Cone-shaped motion vector space (MVS) | Broadcast image | Tactics analysis<br>Player action recognition |
| Pallavi et al. (2008a) | Visual | Weighted graph | Broadcast image | Multiplayer detection & Tracking<br>Solving occlusion problem |
| Pallavi et al. (2008b) | Audio, visual | Directed weighted graph<br>Static and dynamic features | Broadcast image | Ball detection |
| Joo and Chellappa (2007) | Visual | Multiple-hypothesis tracking (MHT)<br>Kalman filter | Broadcast image | Ball and Multi-object tracking |

**Table 3** Continued

| Reference (year) | Data sources | Methodology | Image stream | Application view |
|---|---|---|---|---|
| Jiang et al. (2007) | Visual | Midline detection | Broadcast image | Match statistics |
| | | Global motion analysis | | |
| | | Finite state machine based | | |
| Liu et al. (2006) | Visual and 3D features | Particle filter based on support vector regression<br>Kalman filter | Broadcast image | 3D location of ball and player and tracking |
| Figueroa et al. (2006) | Visual | Weighted graph representation | Four fixed cameras | Player tracking |
| | | Image segmentation | | Match statistics |
| | | Morphological operators | | Solving occlusion problem |
| | | Kinematical motion analysis | | |
| Yu et al. (2006) | Visual | Trajectory-based algorithm | Broadcast image | Ball detection and tracking |
| Tong et al. (2004) | Visual | Condensation algorithm | Broadcast image | Ball detection and tracking |
| | | Region optimization | | |
| Yu et al. (2003) | Visual | Trajectory-based algorithm | Broadcast image | Ball detection and tracking |
| | | Kalman filter | | |
| Yoon et al. (2002) | Visual | Dominant color | Broadcast image | Player detection and tracking |
| | | Mosaicking sequences analysis | | |

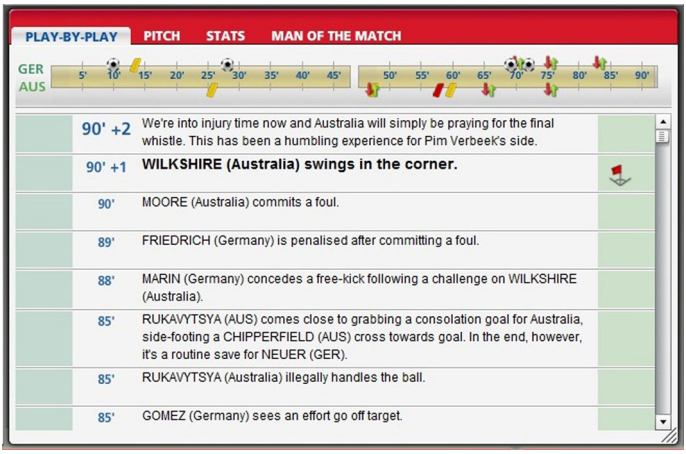


**Fig. 7** Webcast text from FIFA.com

content-based approaches for event detection, results in motivation of using external sources related to sport videos for semantic analysis assistance. Applying ontology-based approaches helps us to process the information manually created for the video and align this information with video according to a specific domain.

### 5.1 Using text sources for soccer video analysis

Two textual sources applied for sport event extraction are closed captions and game log. The closed captions are textual lines which are inserted on videos and provide complementary information in addition to the events occurring in the video for the viewers. Game log is usually entered manually by the operators and describes the important match events as well as their time label, effective person(s), location and result. Assisting video analysis by using text sources decreases the semantic gap between low level and high level features.

In most of the cases, a software application is developed which provides graphical interface for the user to create the text related to each event instantly with accuracy and publish it for the media (Yu et al. 2009). If this text is published through different websites, it is called webcast text and the users can receive the live soccer events by accessing these websites. The available webcast text can be fetched by a web crawler to an intermediate file (Bayar 2010). A sample of webcast text extracted from FIFA website is displayed in Fig. 7. The work in Zhu et al. (2009) has presented a system for attack event detection by analysis and alignment of long shot images with webcast text. The tactical patterns are extracted using ball and player trajectory from broadcast videos and the non-goal attacks are also extracted by means of webcast text information. The attack event is detected by multimodal approaches and based on analysis and alignment of webcast text with broadcast video in semantic level.

Usually a textual event is extracted from webcast text and is detected by using ontology-based methods as a structured text, so the time label which represents the time of event occurrence is obtained. Then the time of match is recognized from the video and the moment in which the event has occurred is obtained by relation of the time label to the corresponding time in the video. The whole event sequence is detected from the video by identifying the video shots and finite state machine modeling. All attack events are extracted using this method and the long shots are aggregated for tactical analysis from detected events. In case

the time info of the game log is imprecise, still the events can be detected by applying specific rules for shots, defining a search range and using a voting mechanism for the rules (Bayar 2010).

Also Xu et al. (2008a) has used the same approach for semantic event detection. For this purpose, first the webcast text is clustered and the textual events are extracted. This task is performed using Probabilistic Latent Semantic Analysis (PLSA). The text and video are aligned together by determining the event occurrence time and event boundaries and by using Conditional Random Field Model (CRFM). Using webcast text along with sport video analysis eases the video semantic detection process significantly. According to this research, events of corner, shoot, foul, free kick, offside, goal and substitution are detected for soccer using the said method. The webcast texts were extracted from Yahoo and ESPN websites. By the same approach in Hu (2010) video adaptation and summary representation under different client's constraints are performed based on semantic utility of each summary, independent of training data. The MMKP (Multi-choice Multi-dimensional Knapsack Problem) was solved for modeling personalized adaptation summary. Goal, shot, free-kick, corner and card events were detected for presentation to the client.

As said before, game log is a text file in which the important events of a match are described. Usually game log is typed by an operator which is time consuming and imprecise for live matches. Therefore, semantic detection from game log and insertion of important events in the video would be impossible. In another hand, for creation of semantic game log in Yu et al. (2009) four operators located in four sides of the field, generate the game log using a software tool. This tool assists the operators in a way that they only responds to the questions such as "when", "what", "where", "who" and "how" by doing some clicks on the graphical interface. The operator submits all the events related to players, referees and soccer video director instantly by using the said questions and the semantic game log are created. Game log is used for important event alert from one hand and from another hand for detecting the video clip boundaries containing that event. However in this research, there is no feature detection necessary thanks to the semantic game log information. The Fig. 8 presents a semantic detection system based on game log.

In addition to important event detection, game log can be used for team tactical information extraction. In this respect, Abreu et al. (2010) presents an automatic tool which is able to implement many tactic analysis viewpoints. The statistics related to goals, pass (successful, unsuccessful), shoot (on-target, off-target) and offside are extractable by this method. The Robocop game log of the years 2007 and 2009 is used instead of match events identification in the proposed system. The simulator available in Robocop soccer league, functions based on a client-server architecture. Consequently each agent connects to a server for playing a game and logs the information related to the played game in the server. At the end of the match, server generates a log file that is processed by the system while its information is extracted from a structured form. The system uses this information and after a sequential analysis process, presents a set of statistics based on defined heuristic rules.

### 5.2 Types of video sources

Using various video sources play an important role in performance and precision of sport video analysis. View angle, image resolution, quality, lighting and frames per second are different among video sources and each source shall be used depending on different applications for sport video analysis and semantic detection. In general, two major sources of broadcast videos and images captured by special cameras are used in sport video analysis domain.

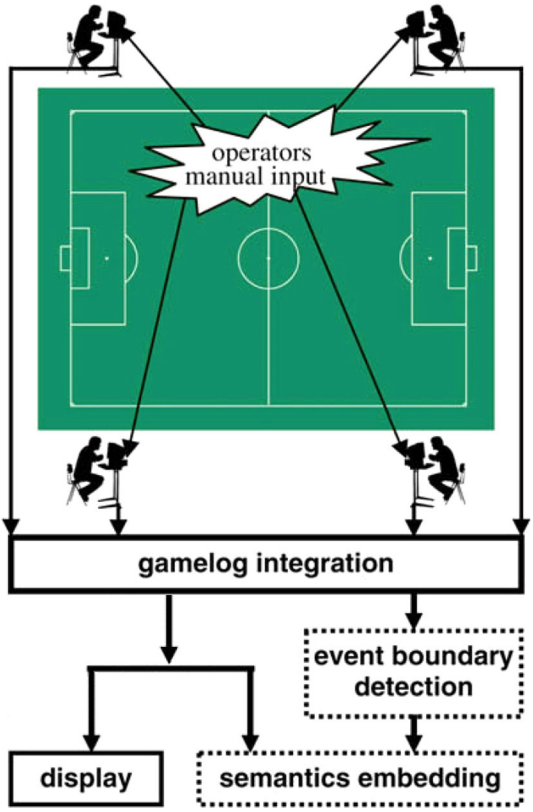


**Fig. 8** A schema of a soccer analysis system based on game log

*5.2.1 Broadcast videos*

During the high class soccer matches, the match scenes are captured, often live broadcast is performed and the video is archived finally. Video shooting is done by the cameras installed at standard locations of the stadium. The location of these cameras is determined as to cover the whole match scenes. The scenes recorded by the cameras are mixed with each other by the TV director based on defined principles and a unit video is established and prepared for broadcasting. In the literature, the broadcast videos being processed are captured using computer TV card. The advantage of broadcast videos is that they are usually accessible for any desired match and easy to archive. However some of these videos do not have sufficient quality and readability due to capturing conditions. Besides, long shots are not suitable for the cases in which more details of the image such as ball and player shape or size is necessary. Also due to the fact that the camera views and the captured images from each angle are limited, they do not have enough performance for solving problems such as occlusion.

*5.2.2 Videos captured by special cameras*

In some cases, broadcast videos are not sufficient for semantic analysis. For obtaining the desired precision, application of arbitrary cameras is inevitable and one or more cameras with special features shall be installed in specific locations to assist the detection of desired events.

Special features are about recording with more frames per second, wide angle recording capability for covering the whole field and zooming capability. The location of the cameras can also be determined for detection of a special event. Making use of multiple integrated cameras enables us to cover the different views and obtain the sufficient details for feature extraction, event detection and object tracking.

### 5.3 Video databases

In the research related to soccer video semantic analysis, the experimental video data shall be prepared for testing the research results. In many works such as Xu et al. (2008a), Tjondronegoro and Phoebe Chen (2009), Zhu et al. (2009) recorded videos by TV card are used for different soccer matches. Most of the videos belong to FIFA world cups, UEFA championships and Olympics of different years with 30 fps rate. Some works such as Ren et al. (2009), Du and Piater (2007), Figueroa et al. (2006) have used images of multiple cameras which do not belong to a specific database. In fact the authors have created their necessary database themselves by applying multiple cameras. For instance, the method proposed by D'Orazio et al. (2009b) has been experimented on videos of Italian championships of the years 2006–2007. DALSA Pantera SA 2M30 cameras are installed for he proposed system in Friuli stadium and are used for video shooting which have a resolution of $1,920 \times 1,080$ and 25 fps video recording speed.

As explained above, no standard databases is yet been generated for soccer video analysis to be used for benchmarking between different woks. Some of them such as Xu et al. (2009) have tested their work on TRECVID which is a valid database including a large amount of standard news videos and ImageCLEF which is a rich medical database so that their work can be compared with other proposed methods in the future. Gathering a set of recent soccer match videos organized by FIFA according to the standards of this organization can be considered as a future work. This database should include different and sufficient standard scenes for event detection and videos from various weather, lighting and grass conditions and for various shapes of ball and players to be commonly used by different works.

## 6 Trend of soccer video analysis

Video semantic analysis is counted as a method for knowledge acquisition from sequence of frames. This issue has been addressed in academic research and in business class applications, however the later have seen so much improvements that has surpassed the former in so many aspects, while the future trend is progressive for both of them.

### 6.1 Knowledge acquisition

The main objective of sport video analysis is to automatically achieve knowledge by applying computer science, image processing and semantic mining for decreasing the knowledge gaps in the sport videos and especially soccer databases. Knowledge discovery is a concept of the field of computer science that describes the process of automatically searching large volumes of data for patterns that can be considered as knowledge about the data.The knowledge obtained through the process may become additional data that can be used for further usage and discovery.

Knowledge acquisition is indeed described as deriving knowledge from the input data. That can be categorized according to (1) what kind of data is searched; and (2) in what form

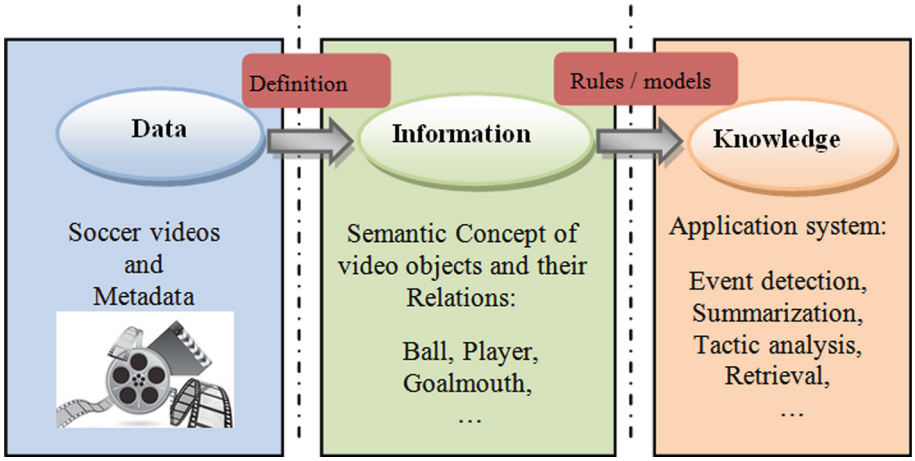


**Fig. 9** Knowledge acquisition for soccer video analysis

is the result of the search represented. Knowledge discovery is closely related to it both in terms of methodology and terminology. Knowledge discovery is closely related to semantic mining, since existing software artifacts contain enormous business value, key for the evolution of systems. Instead of mining individual data sets

The soccer video processing trend for knowledge acquisition is displayed in Fig. 9. In this trend, first some definitions are applied on raw data according to soccer match standards and the main information is achieved. This information includes the ball, players, goal, etc. and their role in the match. In the next step, for extracting semantic information, learning rules and models are applied on patterns and relations according to the main application of the system. These applications may exist among the summarization systems, analysis systems, retrieval systems and even the detection of important events. Eventually, knowledge can be achieved automatically by using the extracted knowledge from the former steps. Analyzing a special team or player during a championship season and determining the number of goals and attacks for a team in the last played matches are instances of high level knowledge which are so important to soccer coaches and analysts; however they are currently performed manually.

### 6.2 Commercial products in the field of soccer video semantic analysis

Not only the professional managers and coaches are interested in automatic soccer video analysis and retrieval, but also the broadcast TV channels are seeking for these kinds of services in order to present technical and statistical analysis of the match to their viewers along with broadcasting the soccer match. In this respect, few systems are developed and introduced to the market as commercial products by some software companies. However the commercial products in some cases have overtaken academic researches and have prepared analysis with a speed and precision higher than the similar works in the literature.

The commercial products can be studied from three different views: prepared services and output, response time and the amount of human intervention. These products provide services such as video indexing and retrieval, statistics preparation, player and team technical and tactical analysis and match summarization. Systems are identified as online, real-time or offline according to their response time. In addition some products use one or more operators for entering information to the system and performing analysis (manual) and some function

automatically. Meanwhile some of them try to increase system speed and response precision by using special purpose cameras.

Some of these products perform the task of receiving, storing and indexing of soccer videos, providing search and retrieval features. The Focus software developed by Elite Sport Analysis[4] is one of these products which provide a wide range of facilities for coaches and analysts in the fields of soccer, basketball, tennis and other sports as well as preparation of match highlight scenes. This software enables instant access to every moment of a match and the event detection is performed manually by operators. The Mambo Studio software developed by Match Analysis[5] also performs offline video indexing and online generation of videos from desired scenes for broadcasting support, content creation and video archiving. In a research project related to Snoek[6] PHD thesis (Snoek 2005) executed in Amsterdam University, a primary sample of soccer video search engine called Goalgle[7] is designed and developed which works automatically. The task of browsing and retrieval is performed by a web-based interface. This interface enables the user to search among the soccer videos which are previously prepared from soccer matches and automatically analyzed. The search is done based on player names, special events, matches and video texts.

Another group of products, work on player and team behavior analysis during a match. MotionView and MotionClip applications developed by All Sport Systems[8] are focused on personal movement and team behavior analysis respectively. They do not support automatic video analysis but provide an efficient graphical user interface and assist the coaches in respect of personal and team behavior analysis and evaluation. This system can also calculate the ball moving and rotation speed semi-automatically. Focus software provides statistical analysis solutions and extensive facilities for the use of coaches and analysts in order to improve their performance and functionality. Also the match statistics and analysis of a special player is presented for boosting team performance.

Some of these commercial products use a set of special purpose cameras for covering all view angles and performing player recognition and tracking. An application software called Gamezone developed by Stats[9] requires three high-resolution cameras to be installed on a base during the match for full coverage of the field. Also the information related to each player is entered into the system's database by the operator and field lines are manually determined. The images of cameras are mixed together and the location of all moving objects of the field including ball, player and referee is determined. Having this information, detailed statistical and individual data extraction is simplified and some information such as distance covered by players, their average speed and region of play is presented to the users. In addition, graphical maps of players' behavior, statistics, results and sport commentary are available online at product website.

Considering the fact that so many researches are dedicated to multi-object tracking and occlusion problem, and this issue is still known as an important challenge, it is still uncertain to have a system solving this problem by mounting only three cameras on the same location. The Prozone[10] software product has more or less the same functionality and performs the desired analysis on the images recorded by 8–12 IP cameras. This system is able to collect

[4] http://www.elitesportanalysis.com.

[5] http://www.matchanalysis.com.

[6] Cees G. M. Snoek.

[7] http://www.goalgle.com.

[8] http://www.allsportsystems.com.

[9] http://www.stats.com.

[10] http://www.prozonesports.com.

the physical, technical and tactical data of the players and teams and analyze them in format of data, animation and multi-layer graphics and provide it for the managers and coaches. Prozone is capable of gathering performance information of a team before the match and foresee the opponent's tactics.

Another sample of these products which seems to be perfect and powerful was deployed at FIFA website during 2010 Worldcup matches. This system displayed the information related to the live soccer match with so much detail for the visitors. The system seems to provide online detail and statistics of the match in a semi-automatic manner and by employing some operators using the images transmitted by the cameras installed in the stadium. All the match highlights are inserted in time order into a table in form of game log. Ball possession of each team in company with the field locations in which the ball has been appeared the most is displayed in form of a heat map. Also the whole match statistics including number of shoots (on-target, off-target), fouls, ball possession, offsides, corners, yellow and red cards and distance covered by the players, is provided.

According to the system developed by FIFA, we can determine the total number of passes per each player and the rate of their correctness. In addition the percentage of total correct passes of each team is accessible and maximum speed and average speeds of the players are determined. After the match and approximately by 1 h delay, the match summary video including match highlights is created and uploaded on FIFA website.

The main features of the said commercial products are comparable according to the Table 4.

### 6.3 Future works and conclusion

In general the soccer video analysis is still at its childhood and its applications and unsolved problems are extending every day. In most of the research works, algorithms based on video processing are presented with increasable precision and speed and their complexity can be decreased as to identify the soccer events faster and more accurately. Robustness of the algorithms should be increased in a way that adapt with the different field, lighting and image quality conditions and present proper results for feature extraction, boundary detection and team tactics recognition. Also the presented algorithms should have the capability of receiving inputs from different types of audio, video, text, etc. so the combination of the extracted features can be applied for optimized event detection. The algorithms which process the game log text should be adapted with different formats and languages and use a better text-mining mechanism, because text sources are a shortcut to semantic features extraction. Match statistic and team tactics are among the items that should be updated and extended based on latest used tactics by the teams in soccer matches.

Despite the fact that commercial products such as FIFA match cast system (http://www.FIFA.com) are shown to be one step further in the aspect of provided services variety, the future challenge of video analysis is in relation with the ever increasing need of automatic and instant semantic mining. Efforts have to be made in order to decrease the response time of the systems and make them online. Besides, the match summary as well as match statistics and analysis must be provided as fast as possible during the match for the users according to their preferences.

The algorithms and solutions presented in the field of soccer video analysis can be generalized and extended to other sport videos. Videos related to other team sports such as basketball, volleyball, etc. which are similar to soccer on the basis of match elements, can be analyzed with some considerations in the available algorithms and heuristic rules.

**Table 4** List of prominent commercial softwares developed for soccer video analysis

| | Motion suite | Focus X2 and X3 | Mambo studio | Gamezone | Prozone 3 | Goalgle | Worldcup 2010 |
|---|---|---|---|---|---|---|---|
| Producer | All sport systems | Elite sport analysis | Match analysis | Stats | Prozone | Amsterdam university | FIFA |
| Manual or automatic | Manual | Manual | Manual and semi-auto | Semi-auto and auto | Semi-auto and auto | Auto | N/A |
| Response time | Online and offline | Online | Online | Online and offline | Online and offline | Offline | Online |
| Extracted events | Based on user | Based on user | Max of all events (based on system version) | Max of all events (based on system version) | Max of all events (based on system version) | Goal, substitution, red and yellow card | Goal, pass, shoot, foul, offside |
| Special cameras | – | – | – | 3 cameras in special positions | 8–12 special purpose cameras | – | – |

## References


Abreu P, Moura J, Silva DC, Reis LP, Garganta J (2010) Football scientia—an automated tool for professional soccer coaches. IEEE conference on cybernetics and intelligent systems, pp 126–131

Ariki Y, Kubota S, Kumano M (2006) Automatic production system of soccer sports video by digital camera work based on situation recognition. In: Proceedings of 8th IEEE inter symposium on multimedia

Assfalg J, Bertini M, Colombo C, Del Bimbo A, Nunziati W (2003) Semantic annotation of soccer videos: automatic highlights identification. Comput Vis Image Underst 92:285–305

Ballan L, Bertini M, Del Bimbo A, Serra G (2010) Semantic annotation of soccer videos by visual instance clustering and spatial temporal reasoning in ontologies. Multimed Tools Appl 48:313–337

Ballan L, Bertini M, Del Bimbo A, Seidenari L, Serra G (2011) Event detection and recognition for semantic annotation of video. Multimed Tools Appl 51:279–302

Bayar M, Alan OZ, Akpinar S, Sabuncu O, Cicekli NK, Alpaslan FN (2010) Event boundary detection using audio-visual features and web-casting texts with imprecise time information. IEEE international conference on multimedia and expo, pp 578–583

Beetz M, Hoyningen-Huene NV, Bandouch J, Kirchlechner B, Gedikli S, Maldonado A (2006) Camera-based observation of football games for analyzing multi-agent activities. In: Proceedings of the 5th international joint conference on autonomous agents and multiagent systems, pp 42–49

Cheng CC, Hsu CT (2006) Fusion of audio and motion information on HMM-based highlight extraction for baseball games. IEEE Trans Multimed 8:585–599

Chen SC, Shyu ML, Zhang C (2005) Innovative shot boundary detection for video indexing. Video data management and information retrieval, pp 217–236

Choi K, Seo Y (2005) Tracking soccer ball in TV broadcast video. In: Proceedings of international conference of image analysis and processing, pp 661–668

D'Orazio T, Leo M (2010) A review of vision-based systems for soccer video analysis. Pattern Recognit 43:2911–2926

D'Orazio T, Ancona N, Cicirelli G, Nitti M (2002) A ball detection algorithm for real soccer image sequences. International conference on pattern recognition, pp 210–213

D'Orazio T, Guaragnella C, Leo M, Distante A (2004) A new algorithm for ball recognition using circle hough transform and neural classifier. Pattern Recognit 37:393–408

D'Orazio T, Leo M, Spagnolo P, Nitti M, Mosca N (2009a) A visual system for real time detection of goal events during soccer matches. Comput Vis Image Underst 113:622–632

D'Orazio T, Leo M, Spagnolo P, Mazzeo PL, Mosca N, Nitti M, Distante A (2009b) An investigation into the feasibility of real-time soccer offside detection from a multiple camera system. IEEE Trans Circuits Syst Video Technol 19:1804–1818

De Sousa J'unior SF, De A. Ara'ujo A, Menotti D (2011) An overview of automatic event detection in soccer matches. IEEE workshop on applications of computer vision, pp 31–38

Du W, Piater J (2007) Multi-camera people tracking by collaborative particle filters and principal axis-based integration. In: Proceedings of the 8th Asian conference on computer vision, pp 365–374

Du W, Hayet J. B, Piater J, Verly J (2006) Collaborative Multi camera tracking of athletes in team sports. In: Workshop on computer vision based analysis in sport environments, pp 2–13

Duan LY, Xu M, Chua TS, Qi T, Xu CS (2003) A mid-level representation framework for semantic sports video analysis. In: Proceedings of 11th ACM international conference on multimedia, pp 33–44

Ekin A, Takalp AM, Mehrotra R (2003) Automatic soccer video analysis and summarization. IEEE Trans Image Process 12:796–807

Eldib MY, AbouZaid BS, Zawbaa HM, Zahar ME, Saban ME (2009) Soccer video summarization using enhanced logo detection. IEEE Int Conf Image Process 43:45–4348

Figueroa PJ, Leite NJ, Barros RML (2006) Tracking soccer players aiming their kinematical motion analysis. Comput Vis Image Underst 101:122–135

Gao Y, Wang WB, Yong JH, Gu HJ (2009) Dynamic video summarization using two-level redundancy detection. Multimed Tools Appl 42:233–250

Gao X, Niu Zh, Tao D, Li X (2011) Non-goal scene analysis for soccer video. Neurocomputing 74:540–548

Gedikli S, Bandouch J, Hoyningen-Huene N, Kirchlechner B, Beetz M (2007) An adaptive vision system for tracking soccer players from variable camera settings. In: Proceedings of the 5th international conference on computer vision systems, pp 21–24

Gonzales R, Woods R (2008) Digital image processing, 3rd edn. Prentice-Hall, Upper Saddle River, NJ

Hartley R, Zisserman A (2000) Multiple view geometry in computer vision. Cambridge University Press, Cambridge, UK

Hashimoto S, Ozawa S (2006) A system for automatic judgment of offsides in soccer games. In: Proceedings of IEEE international conference on multimedia and expo, pp 1889–1892

Hossein-Khani J, Soltanian-Zadeh H, Kamarei M, Staadt O (2011) Ball detection with the aim of corner event detection in soccer video. 9th IEEE international symposium on parallel and distributed processing with applications workshops, pp 147–152

Hu Sh (2010) Personalized content adaptation using multimodal highlights of soccer video. Proceedings of the 11th Pacific rim conference on advances in multimedia information processing, pp 537–548

Hu Sh, Jia Y, Tan Sh (2010) Content aware retargeting of soccer video. 2nd international conference on information science and engineering, pp 1–4

Huang CL, Shih HC, Chao CY (2006) Semantic analysis of soccer video using dynamic Bayesian network. IEEE Trans Multimed 8:749–760

Jiang Sh, Huang Q, Gao W (2007) Mining information of Attack-Defense status from soccer video based on scene analysis. IEEE international conference on multimedia and expo, pp 1095–1098

Joo SW, Chellappa R (2007) A multiple-hypothesis approach for multi object visual tracking. IEEE Trans Image Process 16:2849–2854

Kang C, Hwang J, Li NK (2006) Trajectory analysis for soccer players. In: Proceedings of the 6th IEEE international conference on data mining workshops, pp 377–381

Kang YL, Lim JH, Kankanhalli MS, Xu CS, Tian Q (2004) Goal detection in soccer video using audio/visual keywords. In: Proceedings of IEEE international conference on image processing (ICIP), pp 1629–1632

Khatoonabadi HS, Rahmati M (2009) Automatic soccer players tracking in goal scenes by camera motion elimination. Image Vis Comput 27:469–479

Kim HG, Roeber S, Samour A, Sikora T (2005) Detection of goal event in soccer videos. In: Proceedings of storage and retrieval methods and applications for multimedia, pp 317–325

Kolekar MH (2010) Bayesian belief network based broadcast sports video indexing. Multimed Tools Appl 54:27–54

Kolekar MH, Palaniappan K, Sengupta S, Seetharaman G (2009a) Semantic concept mining based on hierarchical event detection for soccer video indexing. J Multimed 4:298–312

Kolekar MH, Palaniappan K, Sengupta S, Seetharaman G (2009b) Event detection and semantic identification using bayesian belief network. Workshop of IEEE 12th international conference on computer vision, Japan, pp 554–561

Leonardi R, Migliorati P, Prandini M (2004) Semantic indexing of soccer audio-visual sequences: a multimodal approach based on controlled Markov chains. IEEE Trans Circuits Syst Video Technol 14:634–643

Liu Y, Liang D, Huang Q, Gao W (2006) Extracting 3D information from broadcast soccer video. Image Vis Comput 24:1146–1162

Liu J, Tong X, Li W, Wang T, Zhang Y, Wang H, Yang B, Sun L, Yang S (2007) Automatic player detection, labeling and tracking in broadcast soccer video. In: Proceedings of British machine vision conference

Liu J, Tong X, Li W, Wang T, Zhang Y, Wang H (2009a) Automatic player detection, labeling and tracking in broadcast soccer video. Pattern Recognit Lett 30:103–113

Liu J, Tong X, Li W, Wang T, Zhang Y, Wang H (2009b) Automatic player detection, labeling and tracking in broadcast soccer video. Pattern Recognit Lett 30:103–113

Masui K, Dao MS, Babaguchi N (2010) Modeling visual information by spatio-temporal patterns to analyze event tactic in sports video. 2nd European workshop on visual information processing, pp 198–203

Misu T, Gohshi S, Izumi Y, Fujita Y, Naemura M (2004) Robust tracking of athletes using multiple features of multiple views. In: Proceedings of international conference in central Europe on computer graphics, visualization and computer vision, pp 285–292

Misu T, Matsui A, Naemura M, Fujii M, Yagi N (2007) Distributed particle filtering for multiocular soccer ball tracking. In: Proceedings of IEEE international conference on acoustic, speech and signal processing, pp 937–940

Miura J, Kubo H (2008) Tracking players in highly complex scenes in broadcast soccer video using a constraint satisfaction approach. In: Proceedings of CIVR

Miura J, Shimawaki T, Sakiyama T, Shirai Y (2009) Ball route estimation under heavy occlusion in broadcast soccer video. Comput Vis Image Underst 113:653–662

Money AG, Agius H (2007) Video summarization: a conceptual framework and survey of the state of the art. J Vis Commun Image Represent 19:121–143

Nguyen VT, Ly NQ (2010) Query events in soccer video using on-screen texts. IEEE RIVF international conference, pp 1–4

Nillius P, Sullivan J, Carlsson S (2006) Multi target tracking linking identities using Bayesian network inference. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 2187–2194

Nitta N, Takahashi Y, Babaguchi N (2009) Automatic personalized video abstraction for sports videos using metadata. Multimed Tools Appl 41:1–25

Otsuka I, Nakane K, Divakaran A, Hatanaka K, Ogawa M (2005) A highlight scene detection and video summarization system using audio feature for a personal video recorder. IEEE Trans Consumer Electron 51:112–116

Pallavi V, Mukherjee J, Majumdar AK, Sural Sh (2008a) Graph-based multiplayer detection and tracking in broadcast soccer videos. IEEE Trans Multimed 10:794–805

Pallavi V, Mukherjee J, Majumdar AK, Sural Sh (2008b) Ball detection from broadcast soccer videos using static and dynamic features. J Vis Commun Image Represent 19:426–436

Pan H, Van Beek P, Sezan MI (2001) Detection of slow-motion replay segments in sports video for highlights generation. In: Proceedings of IEEE international conference on acoustics, speech and signal processing, pp 1649–1652

Pei C, Gao L, Yang S, Hou C (2009) A ROI detection model for soccer video on small display. In: Proceedings of 3rd international symposium on intelligent information technology application, pp 392–395

Ping Sh, Qing YX (2009) Goal event detection in soccer videos using multi-clues detection rules. 9th international conference on management and service science, pp 1–4

Poppe Ch, Bruyne SD, Walle RVD (2010) Generic architecture for event detection in broadcast sports video. ACM AIEM Proc 10:51–56

Poppe Ch, Bruyne SD, Verstockt S, Van de Walle R (2010) Multi-camera analysis of soccer sequences. 17th IEEE international conference on advanced video and signal based surveillance, pp 26–31

Qian X, Wang H, Liu G, Hou X (2010) A novel approach for soccer video summarization. 2nd international conference on multimedia and information technology, pp 138–141

Qian X, Wang H, Liu G, Hou X (2011) HMM based soccer video event detection using enhanced mid-level semantic. Multimed Tools Appl 55:1–23

Ren J, Orwell J, Jones G, Xu M (2008) Real-time modeling of 3-d soccer ball trajectories from multiple fixed cameras. IEEE Trans Circuits Syst Video Technol 18:350–362

Ren J, Orwell J, Jones GA, Xu M (2009) Tracking the soccer ball using multiple fixed cameras. Comput Vis Image Underst 113:633–642

Ren J, Xu M, Orwell J, Jones GA (2010) Multi-camera video surveillance for real-time analysis and reconstruction of soccer games. Mach Vis Appl 21:855–863

Shen J, Tao D, Li X (2008) Modality mixture projections for semantic video event detection. IEEE Trans Circuits Syst Video Technol 18:1587–1596

Shyu ML, Xie Z, Chen M, Chen Sh. Ch (2008) Video semantic event/concept detection using a subspace-based multimedia data mining framework. IEEE Trans Multimed 10:252–259

Seo K, Ko J, Ahn I, Kim Ch (2007) An intelligent display scheme of soccer video on mobile devices. IEEE Trans Circuits Syst Video Technol 17:1395–1401

Snoek CGM (2005) The authoring metaphor to machine understanding of multimedia. Doctor of Philosophy Thesis

Spagnolo P, Mazzeo PL, Leo M, D'Orazio T (2007) Unsupervised algorithms for segmentation and clustering applied to soccer players classification. In: Proceedings of the international conference on signal processing and multi-media applications, pp 129–134

Sullivan J, Carlsson S (2006) Tracking and labeling of interacting multiple targets. In: Proceedings of 9th European conference on computer vision, pp 619–632

Sun L, Liu G (2009) Field lines and players detection and recognition in soccer video. IEEE international conference on acoustics, speech and signal processing, pp 1237–1240

Taki T, Hasegawa J, Fukumura T (1996) Development of motion analysis system for quantitative evaluation of teamwork in soccer games. International conference of image processing, pp 815–818

Tjondronegoro DW, Chen YP, Pham B (2003) Classification of self-consumable highlights for soccer video summaries. In: Proceedings of IEEE international conference on multimedia and expo, pp 579–582

Tjondronegoro DW, Phoebe Chen YP (2009) Knowledge-discounted event detection in sports video. IEEE Trans Syst Man Cybern 40:1009–1024

Tong X, Lu H, Liu Q (2004) An effective and fast soccer ball detection and tracking method. International conference on pattern recognition, pp 795–798

Vandenbroucke N, Macaire L, Postaire JG (2003) Color image segmentation by pixel classification in an adapted hybrid color space: Application to soccer image analysis. Comput Vis Image Underst 90: 190–216

Wang J, Xu Ch, Chng E, Tian Q (2004) Sports highlight detection from keyword sequences using HMM. In: Proceedings IEEE ICME 27–30

Wang F, Ma YF, Zhang HJ, Li JT (2005a) A generic framework for semantic sports video analysis using dynamic Bayesian networks. In: Proceedings of the 11th international multimedia modelling conference, Melbourne, Australia, pp 115–122

Wang F, Ma YF, Zhang HJ, Li JT (2005b) A generic framework for semantic sports video analysis using dynamic Bayesian networks. In: Proceedings of the 11th international multimedia modelling conference, Melbourne, Australia, pp 115–122

Wickramaratna K, Chen M, Chen Sh.Ch, Shyu ML (2005) Neural network based framework for goal event detection in soccer videos. In: Proceedings of seventh IEEE inter symposium on multimedia, pp 21–28

Xie L, Chang SF, Divakaran A, Sun H (2003) Unsupervised discovery of multilevel statistical video structures using hierarchical hidden Markov models. IEEE international conference on multimedia and expo, pp 29–32

Xie L, Xu P, Chang SF, Divakaran A, Sun H (2004) Structure analysis of soccer video with domain knowledge and hidden Markov models. Pattern Recognit Lett 25:767–775

Xie1 Z, Shyu M.L, Chen Sh.Ch (2007) Video event detection with combined distance-based and rule based data mining techniques. IEEE international conference on multimedia and expo, pp 2026–2029

Xu Z, Shi P (2005) Segmentation of player and team discrimination in soccer video. In: Proceedings of the IEEE international workshop on VLSI design and video technology, pp 369–372

Xu W, Yi Y (2011) A robust replay detection algorithm for soccer video. IEEE Signal Process Lett 18:509–512

Xu M, Orwell J, Lowey L, Thirde D (2005) Architecture and algorithms for tracking football players with multiple cameras. IEEE Proc Vis Image Signal Process 152:232–241

Xu Ch, Zhang Y. F, Zhu G, Rui Y, Lu H, Huang Q (2008a) Using webcast text for semantic event detection in broadcast sports video. IEEE Trans Multimed 10:1342–1355

Xu Ch, Wang J, Lu H, Zhang Y (2008b) A novel framework for semantic annotation and personalized retrieval of sports video. IEEE Trans Multimed 10:421–436

Xu Ch, Cheng J, Zhang Y, Zhang Y, Lu H (2009) Sports video analysis: semantics extraction editorial content creation and adaptation. J Multimed 4:69–79

Yang YQ, Lu YD, Chen W (2004) A framework for automatic detection of soccer goal event based on cinematic template. In: Proceedings of 2004 international conference on machine learning and cybernetics, pp 3759–3764

Yang Y, Lin S, Zhang Y, Tang S (2007) Highlights extraction in soccer videos based on goal-mouth detection. ISSPA, pp 1–4

Yilmaz A, Javed O, Shah M (2006) Object tracking: a survey. ACM Comput Surv 38:13–58

Yoon HS, Bae YLJ, Yang YK (2002) A soccer image sequence mosaicing and analysis method using line and advertisement board detection. ETRI J 24:443–454

Yong LH, Tingting H (2009) Integrating multiple feature fusion for semantic event detection in soccer video. International joint conference on artificial intelligence, pp 128–131

Yu X, Li L, Leong HW (2009) Interactive broadcast services for live soccer video based on instant semantics acquisition. J Vis Commun Image Represent 20:117–130

Yu X, Leong HW, Xu C, Tian Q (2006) Trajectory-based ball detection and tracking in broadcast soccer video. IEEE Trans Multimed 8:1164–1178

Yu X, Xu C, Leong H. W, Tian Q, Wan K. W (2003) Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video. ACM conference on multimedia, pp 11–20

Zhu G, Xu C, Zhang Y, Huang Q, Lu H (2008) Event tactic analysis based on player and ball trajectory in broadcast video. In: Proceedings of conference on image and video retrieval, pp 515–524

Zhu G, Huang Q, Xu C, Rui Y, Jiang S, Gao W, Yao H (2007) Trajectory based event tactics analysis in broadcast sports video. In: Proceedings of the 15th international conference on multimedia, pp 58–67

Zhu G, Xu Ch, Huang Q, Rui Y, Jiang Sh, Gao W, Yao H (2009) Event tactic analysis based on broadcast sports video. IEEE Trans Multimed 11:49–67