Pattern Recognition 43 (2010) 2911–2926

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A review of vision-based systems for soccer video analysis

T. D'Orazio *, M. Leo

Institute of Intelligent Systems for Automation - C.N.R., Via Amendola 122/D-I, 70126 Bari, Italy

ARTICLE INFO

Article history:
Received 30 March 2009
Received in revised form
2 February 2010
Accepted 9 March 2010

Keywords:
Soccer video analysis
Computer vision
Low-level and high-level feature extraction
Modeling of feature dynamics

ABSTRACT

This paper presents a survey of soccer video analysis systems for different applications: video summarization, provision of augmented information, high-level analysis. Computer vision techniques have been adapted to be applicable in the challenging soccer context. Different semantic levels of interpretation are required according to the complexity of the corresponding applications. For each application area we analyze the computer vision methodologies, their strengths and weaknesses and we investigate whether these approaches can be applied to extensive and real time soccer video analysis.

1. Introduction

Soccer is one of the most popular team sports in the world because of the relative simplicity of the rules and the small amount of equipment needed. As a consequence of this popularity, soccer video analysis has attracted much research in the last decade and a wide spectrum of possible applications have been considered such as verification of referee decision, tactics analysis, automatic highlight identification, video annotation and browsing, content based video compression, automatic summarization of play, customized advertisement insertion, graphical object overlapping for better enjoyment of events, player and team statistic evaluations, etc. In some ways, the soccer context can be considered as a specific application of surveillance systems. The methodologies used in surveillance tasks cannot be directly applied in the soccer context since a lot of constraints have to be considered: high variability in the lighting conditions, quick dynamic events, real time analysis, complex situations of occlusions, precise positioning of players in the field, and so on. Algorithms for people tracking, object detection, activity analysis have to face difficult situations such as the overlapping of players wearing the same uniform, unpredictable trajectories, ball not always visible due to difficult lighting condition and a wide camera view, complex events that depend on the positions and the interactions between ball and players. It is for these reasons that the soccer context is very challenging for the scientific community and many papers are going to be published on this subject.

Generally, TV broadcast cameras or proprietary fixed cameras around the playing field are the two possible sources of soccer image streams. Many papers work on broadcast images with the aim of recognizing significant events for television, mobile phone and internet services. On the other side, some works use proprietary cameras suitably placed around the play-field for more specific tasks, not solvable with broadcast cameras, such as 2D/3D reconstruction and visualization of player actions, recording and analysis of team strategies, and evaluation of player performances. Among these works few of them present results for real time processing of soccer images either with broadcast or with proprietary cameras.

In this paper we review the recent developments in soccer video analysis, focusing our attention on the computer vision techniques that have been used to carry out the different tasks for which these systems are devised. In general, the processing framework of soccer video analysis requires different semantic levels of interpretation according to the complexity of the corresponding application (see Fig. 1). There are three major application areas of soccer video analysis: video summarization, provision of augmented information and high-level analysis. In the context of video summarization the main task is the extraction of the most interesting parts of the game (such as goal events, card events, corners, penalties, and so on) to provide for users the ability to access highlights according to their own preferences and to skip the less interesting parts of the video. In this context it is necessary to extract low-level visual features and to model their evolution during the considered events. For the provision of augmented information a greater semantic interpretation of the scene is necessary in order to extract data that can be used by broadcast television to improve their services providing automatically enhanced information. In this case, more complex analysis is required such as the detection of the names of

* Corresponding author.
E-mail address: dorazio@ba.issia.cnr.it (T. D'Orazio).

0031-3203/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2010.03.009

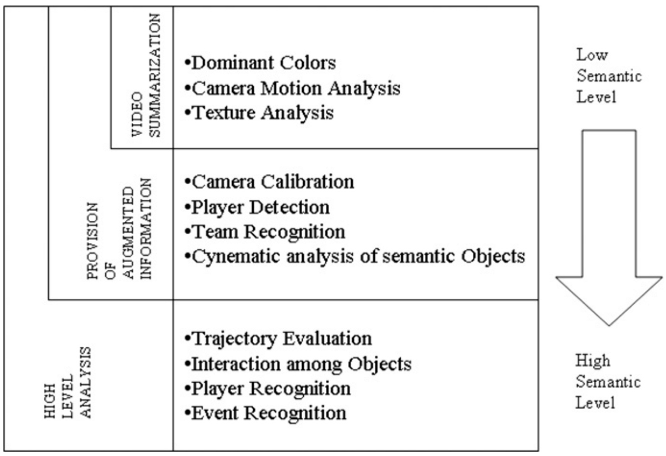

Fig. 1. Overview of different applications of soccer video analysis.

the players involved in the action, the kind of event in which they are involved (penalty, card, goal) and so on. In the high-level analysis more complex evaluations are required for the detection of player skills, the recognition of team strategies, the extraction of tactical formation that produces the scoring opportunities. In this case, accurate localization of positions of players, 3D ball trajectory analysis and action recognition are fundamental steps. A high-level semantic interpretation is required especially when reduced human intervention is supposed and real time computations are necessary for an immediate exploitation of the system output. In this paper we provide an overview of research in the above areas. We collected the most significant papers that highlight the computer vision methodologies used in each application area, and we analyze their key points and their weakness investigating whether these approaches in their present form of implementation can be applied to extensive and real time soccer video analysis systems.

The paper is organized as follow: in Section 2 we review the papers that work on video summarization. Papers on the provision of augmented information are summarized in Section 3. Different aspects of high-level analysis, such as ball tracking, players segmentation and tracking and action recognitions are covered by papers presented in Section 4. In the last section we comment on the novelties and the weakness of the reviewed papers and we discuss the future work that is needed in each application context.

2. Video summarization

As digital sports video data have become more and more pervasive, automatic highlight detection has become particularly demanding. Replacing a long video by a small number of representative segments provides a synthetic description of the video, which can be exploited for numerous applications including both home video and professional usages. Content based-search from large databases, where annotation is too time consuming to achieve manually, requires automatic extraction of the most relevant semantic events. In communication technology contexts, automatic video annotation is important for semantic video transcoding that aims at changing the format of video according to the semantic contents, the communications constraints, and the users requirements. By detecting the most interesting parts of videos, data can be transmitted in different formats and at qualities depending on the application constraints and the users' preferences. Finally, home video applications require the construction of home made video summaries or personalized annotations to realize browsing functions for recorders.

2.1. The computational problem of video summarization

In all the above applications the common problem is to extract from a whole match that lasts 90 min only the most interesting parts such as goal events, card events, corner penalties, and so on. The automatic detection of sequences containing these interesting portions of video could be very difficult if we interpreted the scene from a semantic point of view, but fortunately these videos are characterized by courses of events such as slow motion replay, views of people celebrating, close up of the referee, increase in the noise level of speakers and audience. So, the semantic analysis is generally skipped and this problem is tackled by a feature extraction and a feature evolution modeling.

Video summarization applications are generally off-line and do not impose constraints on the processing time since they can be executed after the matches to annotate the most interesting parts or to extract those videos that should be saved for later analysis.

In particular according to the events that automatic systems have to recognize, different characteristics of the video have to be considered: for goal events, close-up shots of the players, camera shots showing the crowd celebrating, an increase in audio activity, displays of on-screen graphics, or surges in near field motion activities are considered relevant; for card events slow motion replay or short camera view of referee; for corner penalties views of the goal area with many people close to the door, short views of players on the corner of the playing field are considered relevant, and so on. In all these cases, videos can be processed to extract some features such as dominant color, camera motion, image texture, playing field line orientation, change of camera view, text recognition, and so on. Low- and high-level feature extraction techniques can be considered appropriate for this task and well known approaches studied for different application domains [14] can be applied to develop automatic video summarization systems.

Many approaches have been developed for video segmentation and semantic video transcoding in more general application domains which use low-level information to detect semantic events. Some works, applied in different sportive contexts such as baseball, horse racing and so on, use mainly audio features to realize browsing functions for recorders that enable completely automatic detection of sports highlights [15,16]. In this section we will consider those papers which have been applied in the specific context of soccer video analysis and that extract visual features significant for the soccer video summarization.

2.2. Review of video summarization works

In Table 1 we summarize the main works published in this application area. We report for each of them the tasks for which they are developed, the type of image stream, the extracted visual features and the methodologies applied to extract the highlights.

The goal detection task is focused on in different papers [1–4] by the combined analysis of audio and visual features. In the video summarization context this task consists of the extraction of the video sequences starting from the frames in which the actions, that produce a score, begin and ending with the final frames in which players celebrate the goal. In [1] camera motion information and loudness are used, respectively, as visual and audio descriptors. The video signal is processed by extracting low-level visual cues, such as Lack of motion, Fast zoom, Fast pan, whose temporal evolution characterizes different shots. Controlled Markov chains are introduced to describe the feature's evolution during an event of interest. In [2] audio and visual keywords associated with hidden Markov models are used to detect interesting events and extract portions of the video with or

Table 1
A review of video summarization works.

Ref.	Applications	Image stream	Feature extracted	Method
[1]	Goal and shot recognition	MPEG-2 bit stream	Camera motion	Controlled Markov chain
[2]	Break extraction Goal detection	MPEG-1 25 f/s	Camera motion	Hidden Markov model
[3]	Goal detection other significant events	Broadcast	Advertisement detection	Support vector machine
	Other significant events	Sport video	Close-up image detection Crowd image detection Field line orientation	
[4]	Goal detection	Broadcast Soccer video	Pixel and histogram change	Muldimodal analysis
[5]	Play and break recognition	MPEG-1 30 f/s	Dominant color Motion intensity	Hidden Markov model
[6]	Slow motion	MPEG1	Dominant color	Parameter adaptation by pre-game broadcast training
	Goal detection	MPEG7	Kinematic analysis Parallel field line detection	
[7]	Shot detection	MPEG-4	Color, motion, direction and position of individual object	Topological approach
				Spatio-temporal relation
[8]	Play and break detection	MPEG-1 25 f/s	Color and motion features	Dynamic Bayesian networks
[9]	Goal event Corner kick event	MPEG-1 30 f/s	Dominant color Camera motion	Bayesian networks Dynamic Bayesian
[10]	Shot on goal Placed kick Break by offence	Broadcast Video	Color histogram Edge distribution Corner points Camera motion	Support vector machine
[11]	Forward pass Shot on goal Kick off Turnover	PAL stream 25 f/s	Ball motion Playfield detection Player detection	Finite state machine
[12]	Penalty kick Free kick	Hi-vision Camera	Ball detection Player detection	Background subtraction Digital zooming Distance rules
[13]	Free kick, shot Offside, card	Broadcast Video	Text keywords Color, motion Texture features	Webcast text extraction Distance measure Conditional random field

without a goal event. Static visual keywords characterize the possible type of play according to the portion of field view that is covered by the camera. Dynamic visual keywords, based on motion features, describe the camera's motion. A support vector machine in [3] infers the occurrence of events based on a model generated during a training phase that uses audio and visual features. Low- and middle-level image processing techniques, such as crowd image detection, scoreboard graphic location, motion activity measure and field line orientation, are used to detect features. In [4] the authors present a framework for the extraction of soccer goal events in soccer videos by using combined multimodal analysis of audio and visual features and decision tree logic. Pixel change, histogram change together with mean and variance values of background pixels are used to identify grass areas and classify shot types.

Play and break phases are automatically separated in video sequences by the integration of different visual features with motion information [5–9,17,18]. In [5] salient features such as dominant color ratio (see Fig. 2) and motion intensity are selected to model two mutually exclusive states of the game, play and break, using a set of hidden Markov models. Higher-level transitions are taken into account and dynamic programming techniques are used to obtain the maximum likelihood segmentation of the video sequence. In [6], by using kinematic and object-based features, the system can output three types of summary: all slow motion segments in a game, all goals in a game, slow motion segments classified according to objet-based features. The proposed framework includes low-level soccer video processing algorithms, such as dominant color region detection, shot boundary detection and shot classification, as well as higher-level algorithms for goal detection, referee detection and penalty box detection. A content trajectory approach for searching video database is presented in [7] which combines time-varying visual features and time varying visual-spatio-temporal relations between objects to represent and index a shot. Color, motion and position are the visual features that represent individual objects, while relations between objects are modeled by visual features (such as color, speed, direction,…) and spatio temporal relations. A set of color and motion features are employed in [8] in a multimodal multilayer statistical inference framework that recognizes play and break segments using dynamic Bayesian networks. In [17] play or break sequences are recognized by using visual features such as the amount of grass pixels in a frame, the white lines in the goal area or the frame difference for slow motion detection. Kinematic features are used in [18] to locate goal events since they are generally followed by slow motion replays. Goal event, corner kick event, penalty kick event and card event are recognized in [9] by using a semantic analysis based on Bayesian network and dynamic Bayesian network. Video analysis extracts the low-level evidence whereas the semantic analyzer interprets the high-level semantics. The features considered for the inference process are dominant color region, short term motion, texture intensity, logo, parallel lines, score board, black objects, audio energy, and long term static scene.

The knowledge of the soccer domain, the camera work rules, and the broadcast production rules are used in [11–13] for the

Fig. 2. Dominant color ratio as an effective feature in distinguishing different kinds of view in soccer video (from [5]).

extraction of principal highlights in soccer video. The method presented in [11] is suited both for production and posterity logging. The knowledge of the soccer domain is encoded in a set of finite state machines. Visual cues are exploited such as ball motion, play-field zones, player detection and localization. In [12] digital zooming is carried out by automatically recognizing the game situations or events such as when a penalty kick or a free kick is presented. The method is based on player detection and ball tracking by using background subtraction on image sequences captured by fixed Hi-vision cameras. Free kick and goal kick events are recognized using both the coordinates of ball–players and the rules that regulate the changes of the camera work (loose, middle, tight shots) based on the game situations. In [13] webcast text events are synchronized to the events which occurred in the video by the recognition of the salient time tags from the text and the detection of the event boundary in the video. The knowledge of the general broadcast production rules allows the system to model the temporal event structure and to detect the event boundary using a conditional random field model.

The problem of intelligent display of soccer video on small multimedia mobile devices has been considered in [10,19–21]. In [10] a system to generate highlight summaries oriented for mobile applications is introduced and shoot on goal, placed kick and break by offence events are modeled by SVM classifiers trained on simple low-level features. The methods proposed in [19,20] are both based on domain based approaches that consider ground color segmentation, shot classification and determination of regions of interest. In [21] a system that performs automatic annotation of soccer sport video highlights and applies different coding strategies to different parts of the video according to their relative importance for the end user is presented.

In the video summarization applications simple visual features are enough for the highlight detection tasks. The separation of the camera pan during play from the close up view of players is mainly performed using the extraction of motion intensity information in [1,2,5–10]. The detection of the dominant color of the scene, such as the grass color, indicates the appearance of the field and allows the recognition of large camera views (in [6–10,19,21]). The position of the white lines in the images permits the identification of the portion of the field in which the actions happen (in [17]). Also the colors of players/referee's uniforms are used to recognize the close-up views (in [11,12]). Crowd image detection is performed by exploiting the inherent characteristics of such images such as the details and the texture (in [3]). Different solutions have been proposed to model these features: simple heuristic approaches, statistical approaches based on the multiple observations of the events, temporal models that represent the domain knowledge.

3. Provision of augmented information

Among the multimedia applications which are in great demand because of their potential market impact are those for the provision of enhanced content, statistics, dynamic interactive content, advertisements and real time wager placement.

3.1. The computational problem of provision of augmented information

The main problem in providing additional information in automatic processing of video sequences is that the scene content needs to be interpreted and semantic objects need to be correctly extracted. In comparison with video summarization approaches that generally work off-line, approaches for provision of enhanced information have to interpret the scene on-line in order for the information to be made available to users during the sport programs. Of course there is no need of real time processing, since these enhanced contents can be provided also a few seconds or tens of seconds after the events, anyway these systems cannot consider complex image processing techniques.

A deeper scene analysis is necessary for the correct estimation of distances among players, ball and players velocity, multiview calibration and so on. In particular, to carry out these tasks many computer vision methodologies can be considered, such as line extraction for camera calibration, recognition of static playing field features for homography estimation, object detection and tracking for ball and players cinematic analysis. Algorithms applied in other domains for camera calibration such as [22–24], for people detection and tracking [25,26] could in theory be used in this context but the strict constraints of having moving cameras with varying pan, tilt and zoom parameters, low pixel resolution on players in cases of large camera views and real time processing have often limited their applications. The restricted number of works that have been published in this application area, adapt some well known techniques of features extraction, object segmentation and motion detection to the specific soccer context including domain knowledge and manual initialization to complete specific tasks.

3.2. Review of provision of augmented information works

In Table 2 the works on the provision of augmented information are summarized. In this table we report in the columns the application for which they are intended, the image stream, the extracted key features and the processing methodologies. In the last column we report also the prior knowledge that many of these works require in order to carry out the considered tasks.

A few of them [27–29] face the problems of automatic camera calibration and player recognition for the estimation of distances in the playing field, the provision of overlaid data on the game statistics, the forwarding of live event alerts. Some works propose methods for providing users with enhanced images of the playing field such as photo realistic views, virtual flights through real soccer scenes, and mosaic images [30–32].

In [27] the player segmentation problem is performed by a region based analysis and retrieval which uses prior knowledge such as the approximate color of the uniform, the shorts and the

Table 2
A review of works on provision of augmented information.

Ref.	Applications	Image stream	Feature extracted	Method	Prior knowledge
[27]	Player segmentation for augmented information	Broadcast images	Statistical properties of region	Adjacency graph matching	Knowledge of color uniform
[28]	Graphic overlaid using camera pose data	Standard broadcast images	Line and arc detection	Spatialized Hough transform	Initial estimation of camera position
[29]	Live event alert instant semantic embedding	TV broadcast images	Dominant color lines detection	Multiboundary event detection	Manual input of gamelog
[30]	Photo realistic image generation from virtual viewpoints	Video images from fixed cameras	Interest points extraction of region	Quasi-dense depth map generation Motion segmentation View interpolation	Manual segmentation to reduce artifacts
[31]	Virtual view synthesis	Fixed overlapping cameras	Dynamic region extraction	View interpolation	Manual selection of corresponding points
[32]	Image rectification	TV images	Lines and points tracking	Image to image homography	
[33]	Video content analysis	Raw unedited broadcast video	Field lines goalmouth detection	SVM classifier for event moment detection	

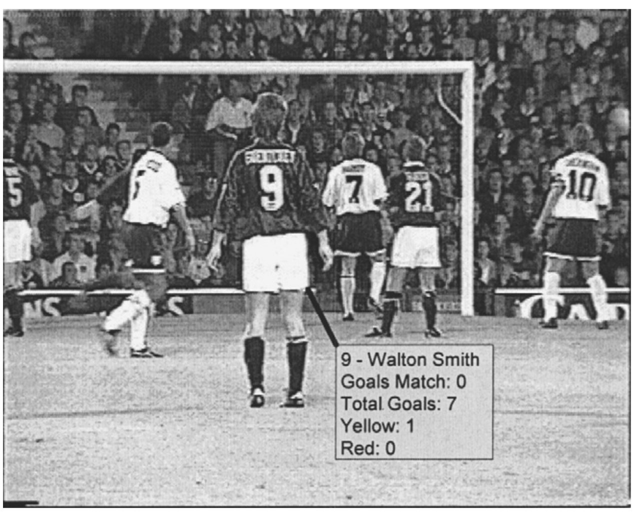

Fig. 3. Augmented information using overlay (from [27]).

color of the grass. Regions are selected based on a statistical description, and interconnecting arcs between them, allow the definition of a topological graph description that represents the object signature. Graph matching is applied for tracking using a fixed descriptor (constructed manually) or an automatic descriptor (generated by defining a bounding box in an initial frame). Augmented information can be supplied to users in the form of overlays as shown in Fig. 3. In order to be able to place virtual annotations on the playing field, to indicate things such as distances between players and the goal, or whether a player is offside it is necessary that the camera positioning, orientation and focal length be estimated in real time so that the graphics can be rendered to match the camera view. In [28] a method for accurate camera position determination using multiple images is presented. The method uses markings on the playing field, such as arcs and lines, to compute the camera pose. The authors also propose a means for automatically initializing the tracking process which makes use of a modified Hough transform. Example of some typical overlay data produced by this approach are shown in Fig. 4. A semi-automatic instant semantics generation system for live soccer video is presented in [29] for two possible applications: live event alert and instant semantics embedding. To quickly and accurately acquire the semantics of a live soccer video, the approach allows easy manual input of the gamelog, combines direct input from the referee and the broadcast director of the soccer game, and conducts multiboundary detection of events. Low-level features such as dominant color, center line, goal mouth, are used for the boundary decision of the selected events.

In [30] the authors propose a number of cameras that populate a stadium, and by a joint view interpolation algorithm produce photo realistic moving or static images from virtual viewpoints, where there is no physical camera. A quasi-dense disparity map is constructed starting from the extraction of interest points that are used as seeds to propagate the matches in its neighborhood. Then, a method based on a probabilistic model of object occupancy and visibility is implemented to segment and track non-rigid objects in one or more views. A method for virtual view synthesis is presented in [31] that allows viewers to virtually fly through real soccer scenes which are captured by multiple cameras in a stadium. The method generates images from arbitrary viewpoints by the view interpolation of real camera images near the chosen viewpoints. The projective geometry between neighboring cameras is used to synthesize new view images. The method of view interpolation is applied only to dynamic region obtained by subtracting the background from the original image. The multiple video cameras are carefully placed in the field so that the size of players and the overall colors in the captured scene are almost identical across the cameras. In [32] an approach for incrementally estimating the homographic transformation between a model and an image is presented. The estimation of the transformation can be used for image rectification and mosaicing. The image-to-image homography estimation technique is based on the use of interest points combined with straight lines.

In [33] the problem of automatic broadcast soccer video generation is examined. The main aim of this work is to address the time critical and labor intensive operation in a commercial broadcasting situation, producing automatic sport video composition that could help in the task of generating broadcast video for professional directors. The input to the system is the raw soccer videos recorded at the soccer stadium, and the automatic generation is similar to that produced by professional soccer broadcasters. The processing mainly includes two issues: detection of the replay-worthy soccer events from the unedited main camera soccer video and then the generation of replay scenes for the detected events; the analysis of the video content from multiple camera inputs to perform the multicamera view switching.

Applications for provision of augmented information require a more profound study of video contents such as the detection of player positions in the playing field, the recognition of players, the extraction of metric information from multiple cameras and so on. Nevertheless, many of them cannot be considered fully automatic or really suitable for on-line applications since, as the last column

Fig. 4. Graphics overlaid using camera pose data (from [28]).

of Table 2 shows, a strong human intervention is required to provide the prior knowledge needed during the processing, such as the manual selection of corresponding points, the initial camera calibration, the knowledge of the uniform colors and so on.

4. High-level analysis

Due to the increase of computing power over the last decade, the demand for systems capable of performing high-level analysis in sport has resulted in large research efforts. Based on the camera image streams much high-level analysis can be done to answer questions such as: What are the characteristic offensive plays of the two teams? What are the strengths and weakness of a particular team player? What roles do the players have? Do their capabilities match their roles? Do they achieve their tasks? How does a team create scoring opportunities? What are each players' skills? What is the tactical formation of a team?

At the same time the analysis of real time events is receiving particular attention from referee associations, the sports press and supporters. During soccer matches a number of doubtful situations arise that cannot be easily judged by the referee committee. Automatic visual systems objectively checking image sequences would prevent wrong interpretations due to perspective errors, occlusions, or the high velocity of events.

4.1. The computational problem of high-level analysis

For answering such questions and addressing the above tasks, systems are required to have the capability to estimate reliably and accurately the positions of the players and the ball, and of generating a compact representation of motions and their segmentation into actions.

The common characteristic is the need to detect and track moving objects relevant to the considered analysis. For goal event detection it is necessary to evaluate the ball position with respect to the goal line plane to detect if a goal has been scored. In more complex analyses for coaches, tacticians, and computer assisted refereeing, the players and the referee positions in the field together with the ball tracking provide valuable information. For other crucial events, such as the detection of infringements, such as a player touching the ball with his hand during a particular play, it is not only necessary to perform ball and player tracking but also to provide an accurate analysis of the player's posture and interaction with the ball.

According to the applications, high-level analysis systems could impose also strict real-time processing and reliability constraints: computer assisted refereeing systems have to work in real time and also provide valuable estimations in terms of accuracy; player and team statistical analysis could be very useful for coaches and tacticians if they were provided on line during the match to modify and improve the game strategies.

The main tasks that have to be addressed in all the above applications concern fundamentally three main areas: (1) single object detection, recognition and tracking, (2) multipeople detection and tracking, and (3) multiview trajectory analysis. In particular these areas are strictly related to other research fields concerning as well camera calibration, motion detection, people recognition, shape representation, semantic event analysis and so on.

Many recent algorithms for background subtraction [34,35], single camera [36] and multicamera people detection and tracking [37–39], trajectory analysis [40] developed in the last years in different domains can be used to solve some of the above tasks. Actually many well-known methodologies [41–43] have been the starting points of papers that fall in the category of high-level analysis, but as autonomy and real time requirements increase, details and object resolutions decrease, either more complex methodologies and more semantic object interpretations have been introduced or domain knowledge and manual model initialization have been considered.

4.2. Review of high-level analysis works

In the following subsections we review the papers according to the key points they focus on: we start from papers that work on ball detection and tracking, then we analyze systems that segment and track players; some papers that collect the information and use statistical analysis are considered, and finally a few papers on real time event detection are described.

4.2.1. Ball detection and tracking

In soccer matches the ball is certainly the focus of attention for the audience since it generates all the most significant events. In Table 3 the works on ball detection and tracking are summarized. In this table we report in the first columns the focus of each work, the image stream, the key points, and in the last two columns the test sequences on which the experiments were carried out with the computational costs.

The problem of ball detection is very difficult when images are taken from fixed or broadcast cameras with a wide camera view since the ball is represented by a small number of pixels and the images are of different sizes, shapes and colors (see Fig. 5). In these cases, many works propose approaches based on the evaluation of the ball trajectory [44–48,51] since the analysis of kinematic parameters allows the ball detected from among a set of ball candidates.

Direct detection algorithms based on pattern recognition approaches work well only with zoomed images because the ball texture is visible, and where there are not large variations in the appearance of the ball [50].

The problem of ball detection in broadcast soccer videos has been faced in [45–49]. In [45] a trajectory based detection and tracking algorithm for locating the ball is presented. In the first phase, a set of ball candidates, selected according to the ball size, color, shape, and consequently a set of ball trajectories is

Table 3
A review of works on ball detection and tracking.

Ref.	Focus	Image stream	Method	Test sequences	Comp. costs
[44]	Multiview tracking eight cameras	Eight fixed cameras 720 × 576	Model-based 3D position estimation	8 seq of 5500 frames	3 s delay
[45]	Single view tracking	Broadcast images MPEG-1	Trajectory based	2085 frames 52437 frames	off-line
[46]	3D ball position by monocular video	Broadcast images	Viterbi algorithm	Synthetic data 2 × 600 real img.	off-line
[47]	Ball route under overlapping	Broadcast images	Generation of ball route candidates	2289 imgs.	Off-line
[48]	Interaction of the ball and players	Broadcast images 960 × 540	Ball Trajectory Evaluation	600 imgs	off-line
[49]	Ball detection	Broadcast images	Ball Motion Analysis	338 frames 12 seq	off-line
[50]	Ball recognition	Fixed camera 532 × 512	Neural network approach	821 imgs	3×10^{-1} s/fr
[51]	3D trajectory	Two fixed cameras	Particle filters	1 seq of ∼ 150 imgs	0.5 s/fr

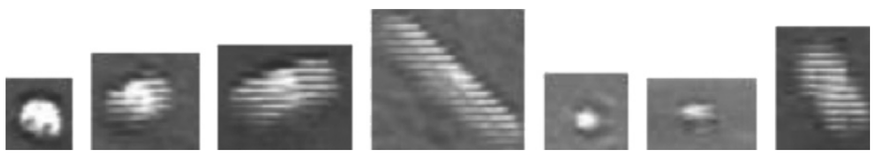

Fig. 5. Different ball samples in various sizes, shapes, and colors (from [44]).

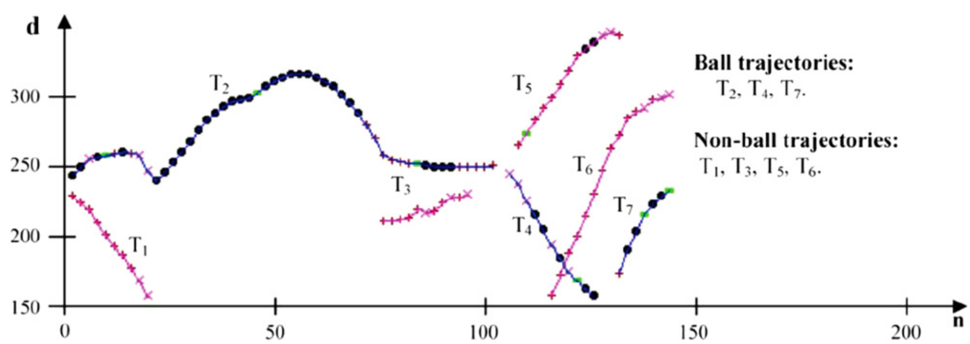

Fig. 6. Illustration of candidate ball trajectories (from [45]).

extracted (see Fig. 6). Then, studying the trajectory information, it is relatively easy to recognize the ball as the most active object among the candidates in the soccer video. Possible false positives, such as parts of T-shirt or socks of players, do not move significantly during the game. Also in [46] several features such as the color, the size, the ratio of the length and the width of the ball shape, are considered to extract ball candidates in each frame, then a Viterbi decoding algorithm is applied to extract the optimal path which is most likely to be the ball's path in consecutive frames. Once the ball is detected the Kalman filter and template matching tracking procedure is started, using the ball location to update the template and to guide possible ball re-direction. The problem of estimating the route of the ball when it overlaps with players and lines in broadcast soccer video has been faced in [47]. The frames between the disappearance and the reappearance of the ball are analyzed to generate a possible ball trajectory for ball route candidates and select the trajectory that is the most probable ball route. The interactions of the ball with players that cause occlusions and overlapping is considered in [48]. In these situations the particle filter tracking is suspended and for each player who is suspected of having e the ball, the ball is searched for in a predetermined area with the player position as the center. After the detection of the reappearance of the ball the trajectory that accumulates the most evidence is selected and ball tracking is resumed. In [49] hybrid techniques are used to identify the ball in medium and long shots. Candidate ball positions are first extracted using shape and size, then motion information is used to filter candidates in medium shots, whereas dynamic programming is applied to search the longest path of weighted graphs constructed for ball candidates in long shots.

In [44] multiple fixed cameras are used for real time modeling of 3D ball trajectories. A motion based thresholding process detects the candidate balls, while a probability measure calculates the likelihood that each moving object represents a ball. Then, the 3D ball motion is modeled as a series of planar curves. Reliable estimates are obtained by triangulating multiple views. Finally, the ball trajectory is modeled as one kind of ball motion, rolling, flying, in-possession, and out of play.

A neural network approach is used in [50] on video images acquired by fixed cameras for the real time detection of the ball in goalpost areas. A set of ball candidates is generated by using a Circular Hough Transform, then different neural networks, trained with different light conditions, are used to select the best candidate to contain the ball (see Fig. 7). The appropriate preprocessing of the ball candidate images is selected in order to guarantee the best performances of the ball recognition routine. Two opposite views of the ball are used to evaluate the 3D ball trajectory and solve occlusion problems. In [51] a distributed state estimation architecture is proposed for 3D ball tracking. Networked subsystems, equipped with a self-contained particle filter, can operate in stand alone as well as in network mode by modeling the dynamics with ballistic motion, bounce and rolling. Ball candidates are extracted by using a hybrid technique of chroma keying and inter frame subtraction.

Fig. 7. An image of the ball acquired by a fixed camera in the goal area (from [50]).

As the last column of Table 3 shows, trajectory based approaches are generally off-line since the evaluation of the kinematic parameters for all the ball candidates requires a long period of observation. On the other side, ball recognition approaches, which work on still frames, impose strict camera views in order to have the ball texture visible and the ball shape not to be deformed.

4.2.2. *Player and referee detection and tracking*

As summarized in Table 4, many works have been presented in the last years on player detection and tracking, covering different aspects from player segmentation and shadow removing with fixed cameras [52,75,53] to player classification and blob merge solution with moving cameras. The player and referee detection problem has generally been faced by segmenting blobs on the field that are not green [55], and applying size constraints and color models [56,57,76]. Player classification and team discrimination have been faced by applying initial training phases during which the models of players are provided in a supervised way by human operators who manually select the regions corresponding to different uniforms [54,77] or in unsupervised way by automatic clustering approaches [60,61,74]. Multiple players in group situations are generally separated by trajectory analysis approaches [63,70,49], by multiple hypothesis approaches [62,65,64] or by multiple cameras systems that allow disambiguation among different player configurations [65,67,69,71,78].

Table 4
A review of works on player detection and tracking.

Ref.	Focus	Image stream	Method	Test sequences	Comp. costs
[52]	Background recovery Shadow removing	Fixed camera	Morphological operations	4500 frames	1.4 f/s
[53]	Shadow detection	Fixed camera	Skeletonization Spatial Filtering	2000 frames	-
[54]	Team discrimination	–	Adapted hybrid color space	Few images	–
[55]	Player Segmentation and tracking	Broadcast images	Grass field extraction Line marks detection	8 min long sequence	off line
[56]	Player segmentation	Broadcast images	Morphological segmentation	Few images	6–7 f/s
[57]	Player	Broadcast images	Relevant color detection	Full match (for a qualitative analysis only) 344 frames	–
[58]	Detection				
[59]	Action Recognition		Camera calibration Motion interpretation detection		
[60]	Player	Broadcast images	Unsupervised learning of player models	$\sim$ 1000 frames	–
[61]	Detection ad labeling		Markov chain data association		
[62]	Multiple target	Fixed cameras	Track graph	10 min video	–
[63]	Tracking and labeling		Trajectory feature extraction		
[64]	Multiple object tracking	VS-PETS dataset	Multiple association hypothesis	600 frames	Real time
[65]	Player tracking in complex scene	Broadcast images	Player transition graph Graph hierarchy	$\sim$ 300 frames	off line
[66]	Player tracking	Broadcast images cameras	Directed graph motion analysis	1300 frames construction	0.27 f/s
	Player				
[67]	Multicamera Tracking	Fixed cameras	Multiview Kalman trackers	2×5000 frames	-
[68]	Multicamera	Two uncalibrated cameras	Belief	200 frames	-
[69]	Tracking		Propagation		
[70]	Tracking Player	Four digital cameras	Kinematic motion analysis	45 min video	–
[71]	Multiview tracking of players	Three fixed cameras	Features matching by multiple cameras	210 frames	0.18 f/s
[72]	Multicamera	TV images	Player models based on interest points	400 frames	6–8 f/s
[73]	Tracking				
[74]	Multicamera player tracking and labeling	Six fixed cameras	Unsupervised clustering of player histograms	Five sequences 2 min long	–

In the columns of Table 4 the focus, the image sequences and the computational costs are reported for the considered works on player detection and tracking.

The problem of background recovery for soccer player segmentation is considered in [52]. A non-parametric morphological leveling operation takes into account the specific problem of lighting changes and the fact that slow and fast motion can be present in the scene. The paper also discusses the problem of shadow reduction and proposes a simple morphological transformation to eliminate them, based on some specific geometrical features (see Fig. 8). In [53] an unsupervised learning procedure is used for shadow classification in order to successfully track individual targets. The RGB colors distribution of the foreground and shadow classes of feature data is generated, and a skeletonisation and spatial filtering process is developed to identify the components in the foreground segmentation that are most likely to belong to each class of feature. The effect of the segmentation results on the estimated trajectories of players was evaluated in [75]. Since players are often motion blurred due to the fast motion of the camera, it not always possible to segment them correctly. The authors propose the use of a K-means algorithm for the accurate segmentation of players' legs. In [55] the player segmentation is carried out by eliminating fast camera motion effects through the correspondence between line marks in soccer field models and the image sequence.

In order to identify the players by the colors of their uniforms, in [54] an adapted hybrid color space is built by means of a sequential supervised feature selection scheme associated with a given family of images. Its dimension is not always equal to three, as for classical color space, but it is specifically designed to yield the best discrimination between the pixel classes. In [77] the authors propose a method for player segmentation and team discrimination based on the detection of a mean distributed color feature to extract the field region and use only hue and saturation components in HSI color space to eliminate the effects of illumination. In [56] color histograms of the athletes' uniform and of the turf areas are built interactively during rehearsal by manually selecting the areas of interest, then morphological segmentation is carried out in two steps: a coarse segmentation by binary reconstruction based on the areas detected by thresholding the color histogram, and then a fine segmentation by watershed transformation with markers.

The visual perception of players and ball is decomposed in [57] in three interrelated subproblems: the identification of relevant color regions, the estimation of the pan and tilt angle of the camera and the zooming parameter, the identification and positioning of the players and the ball that are in the camera view. The player detection procedure uses special templates to calculate likelihood-maps for player locations based on color distributions, compactness of the segmented region and the vertical spacing inside the regions [58]. Some results with real data from live coverage of World Cup 2006 games in Germany are presented in [59] both for camera parameter estimation and player tracking. Color models are learned semi-automatically using K-means clustering on manually marked regions. Player appearance models are learned, by using an unsupervised approach, in [60,61] from hundreds of samples automatically collected by detection (see Fig. 9). A boosted cascade detector is trained on the Haar features extracted from the sample set. Thereafter these models are used for player labeling. In the detection phase playfield segmentation by learning the dominant color (corresponding to the grass color) is first used to filter out background regions.

The problem of multitarget tracking when many targets with indistinct appearances frequently occlude one another is discussed in [62,63]. Initially, in order to distinguish different targets a PDF for the RGB values for each category is learnt from a few labeled training examples. Track graphs, denoting when targets are isolated and describing how they interact, are generated. Then, each target's path is searched through the graph exploiting the constraints imposed by the graph structure. In order to establish the player's identity his position relative to his teammates is analyzed. A similarity measure between tracks is used to infer the most likely configuration of paths for all targets. In [64] a multiple-hypothesis approach handles objects entering and exiting the view, merging and splitting as well as objects that are detected as fragmented parts. The association problem between targets and observation data increases exponentially with the number of targets. The authors propose an efficient methods of maintaining multiple association hypotheses with the highest probabilities over all possible histories of associations. The data association problem is posed as a minimum weight edge cover problem. In [65] a graph based approach is used to cope with difficult cases such as tracking many players in a small area for a corner kick. Graphs of player blobs representing possible player transitions are constructed. The view of each blob provides a constraint on the number of players in the blob and such constraints are propagated through the graph to reduce the ambiguities in the numbers. The remaining ambiguities after the propagation are handled by a statistical approach in which a set of the most likely interpretation of the numbers is selected. Finally the players' trajectories are determined based on their smoothness. The player tracking problem is formulated in [66] as a path

Fig. 8. The player segmentation (from [52]).

Fig. 9. Some positive examples of players used for player model learning (from [60]).

finding problem in a weighted graph where each target object is a node and the weights are indicative of the strength of association between them in two successive frames.

Many works propose multiple camera systems with fixed cameras for player tracking. A system architecture and method for tracking people is presented in [67]. The system input is video data from static cameras with overlapping fields of view at a football stadium. In the first stage the data from single cameras are processed, then in the second stage a multicamera tracker is applied by a Kalman filter to model the player position and velocity. Under the assumption that the camera overlooking the playground is static and positioned such that its optical axis is perpendicular to the floor, in [76] a color based probabilistic tracker is presented. A multiple player tracker is proposed as a set of separate trackers, one for each player, which are combined by inferring a Voronoi partitioning among all players at each time step, and tracking each player within its Voronoi cell. An approach to tracking athletes in team sports using multiple cameras is proposed in [68,69] and addresses several issues including occlusions and propagation of incorrect information. It is based on the use of belief propagation which enables good observations in some views to compensate for poor observations in other views due to occlusions. The observation process consists of matching the color histograms in a set of sampled regions with a previously learned reference model. In the experiments, the authors manually initialize the region of athletes in the first frame acquired from each camera and allow the system to learn the reference models. Four static cameras are proposed in [70] to develop a tracking system that analyzes a whole soccer game. By considering a model of the players and some morphological operations the authors treat occlusion and congestion problems by splitting segmented blobs representing a set of two or more partially occluded players. The splitting process is done using a graph representation in a backward and forward direction. Multiple views are used in [71] to track athletes using multiple visual features. The trajectories of objects in the world coordinates are refined using various features such as color texture, region and motion in the 2D images acquired by the cameras. The trajectory refinement process is done also by weighting them adaptively to their self-evaluated reliability. Experiments were provided on one real professional football sequence, with a processing time for three overlapping views of 5.5 s per frame.

Sixteen cameras are proposed in [78] to cover all the field and to track all players in parallel. In inner camera operations, the player's foot coordinates are calculated and uniform classification applied. Next, in inter camera operations, the information of the 16 cameras is integrated for player tracking. A simple distinction between occluded player and non-occluded player regions is made based on the height and area of player regions, since occluded players must be large. Only non-occluded player regions are considered for further processing.

A modular multicamera framework is presented in [72], designed for rotating and zooming cameras. For each camera an estimate of the homography is maintained and updated as frequently as possible. The system does not use background models but instead takes advantage of a tracking method based on local features presented in [73]. However, the regions of interest, in which the feature selection has to be applied, are selected manually in the initial frames of the sequence. In [74] a method for player tracking and team discrimination was applied on image sequences acquired by fixed cameras distributed on the two opposite sides of the field. When players enter into the filed an unsupervised clustering algorithm is started to search five different classes corresponding to the uniforms of the two teams, the two goal keepers and the referees. Some constraints such as the position and the number of the players, their distribution on the field allow the system to automatically associate the extracted color classes to the appropriate team, goal keeper or referee classes (see Fig. 10).

The large number of works published on player tracking and team discrimination demonstrates that this problem is very difficult and a clear assessment has not been reached. Well known and well assessed methodologies of object tracking cannot be directly applied to player tracking since the initialization phase is not immediate and also there is the objective difficulty of following players with similar uniforms that merge in group blobs and then separate resulting in great ambiguities in the tracking results. As Table 4 shows, many works are off-line, since they require the analysis of all the history of each entity to solve group blobs, or suggest the use of multiple cameras to solve the ambiguities using different views.

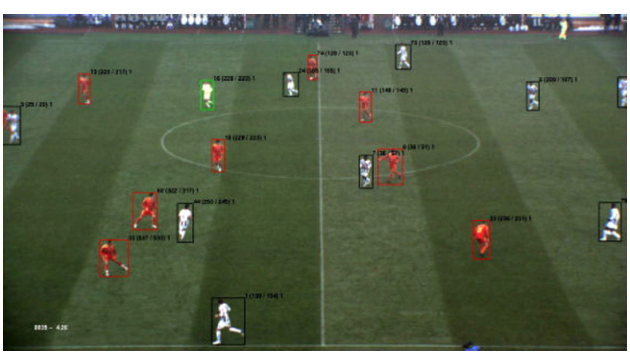

Fig. 10. Output image after the classification phase: boxes of the same colors refer to players classified as belonging to the same team (from [74]).

4.2.3. *Team statistical analysis*

Some works have been presented in literature for tactical and team statistical analysis. In Table 5 we report the focus, the image stream, the key points, and the test sequences for each paper. Starting from the ball and players tracking information collected during the matches, these works try to extract tactical

Table 5
A review of works on team statistical analysis.

Ref.	Focus	Image stream	Method	Test sequences
[79]	Tactical presentation of goal event	Broadcast images	Support vector classification	168 goal events
[80]				
			Particle filter tracking	
[81]	Distances covered by players	Fixed overlapping cameras	Image calibration for image-object transformation	Four matches
[82]	Player possession	Broadcast video	Support vector classification	Half game
[83]	One player performance evaluation	Fixed cameras	Mapping between image and field coordinates	Some real and simulated video clips
[84]	Player trajectory analysis	Football Manager 2005	Relationships between trajectories	Simulation game
[85]	Football game analysis	Robocup simulated matches	Probabilistic action model	Simulated game

information for trainers [79,80], to evaluate player skills [81,82,84] and to perform game analysis [83,85]. These systems are not completely automatic but can greatly improve manual work during post processing analysis.

In [79,80] a system for showing goal events in a tactical mode to the coaches and sports professionals is described. After an initial phase in which goal events are detected by the analysis of web-casting text and broadcast video, tactical representations, known as "aggregate trajectory", are constructed based on multiobject trajectories using the analysis of temporal-spatial interactions among the players and ball. The acquisition of player trajectories in far-view shots is achieved by play-field detection using Gaussian mixture color models and support vector classification on player candidates. A support vector regression particle filter keeps tracking player in the frames. The distance covered by soccer players was measured in [81] with an automatic tracking system that is corrected manually when complex situations, such as when the player's trajectory changes during periods of occlusions, are not solved automatically (see Fig. 11). The segmentation and tracking phases for each game require, respectively, 6 and 4 h of processing and in order to calculate the image-object transformation, before the games, 20 control points were established and measured directly onto the field.

In [82] a semi-automatic system was developed to acquire player-possession for broadcast soccer video, whose objective is to minimize the manual work. To acquire player possession the authors try to recognize the players touching the ball by assuming that they are those closest to it. Support vector machine methods are used to recognize the team of the player touching the ball. The view information and the player roles are used to produce the candidates of the player touching the ball. The selection among the possible candidates is done manually by the operator of the system.

Scout is a system presented in [83] for event game speed analysis and tracking. Background subtraction, connected component labeling, morphological filtering are used to segment and track moving objects. A vanishing point based method is proposed to map between the screen and physical coordinate systems. The system is designed to evaluate one player's performance at a time. In the case of complex player occlusions, more sophisticated tracking modules will be needed to avoid manual intervention. Analyzing the trajectories of moving objects, which consist of 22 players and a ball, in [84] much useful information is extracted in order to evaluate the performance of several players in a quantitative way. The proposed model is based on the trajectories of the players and the ball and their relationship. Several performance measures are introduced to analyze the performance concerning the interactions between the players of the same team and of their adversaries. The experiment was executed by collecting the trajectories of players and ball from a simulated soccer game. In [85] a model-based game analysis is carried out to map the real game process into an abstract representation obtained from features specifically designed for game analysis objectives. The system requires a real time positioning system that can continually track the position of the players' feet and the ball with an accuracy of a few centimeters. The authors propose the usage of Cairos Technologies (RFID-based technology) to collect the data and test their methodology to recognize passes, shots, and dribbles on Robot cup simulated soccer matches.

The autonomy and performances of these systems depend on the preliminary steps of ball detection and player tracking. Failures or uncertainties in these preliminary steps impose the need for human intervention to solve ambiguities or difficult situations. Improvements in the player and ball tracking algorithms will certainly lead to greater interest in developing automatic systems for tactical analysis.

4.2.4. Real time event analysis

The analysis of real time events is getting a lot of attention from referee associations, the sports press and supporters. During

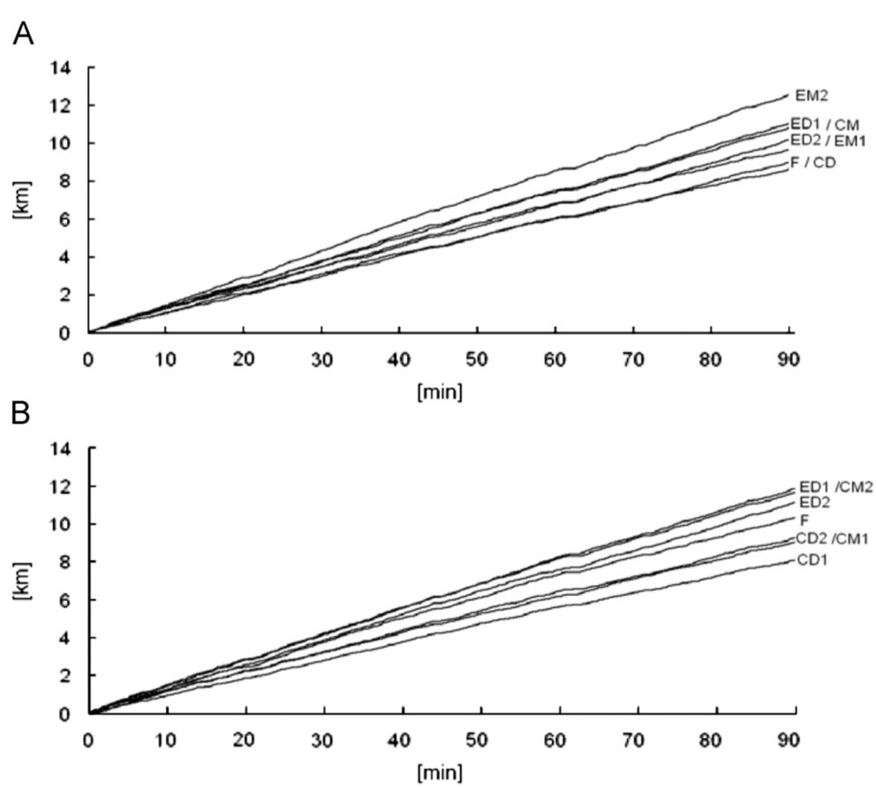

Fig. 11. Distances covered by the outline players of two teams (A and B) of Brazilian soccer players during the 90 min of a game. The players were classified as central defenders (CD), external defenders (ED), central midfield players (CM), external midfield players (EM) and forwards (F) (from [81]).

Table 6
A review of works for real time event analysis.

Ref.	Focus	Image stream	Method	Test sequences	Comp. costs
[86]	Real time goal detection	Four fixed camera	Ball recognition and tracking by neural approaches	Eight matches	200 f/s
[87]	Offside detection	16 fixed cameras	3D ball trajectory Player tracking	Two real sequences	–
[88]	Offside detection	Broadcast image	Field extraction Player classification Ball tracking	Two sequences	4 f/s
[89]	Offside detection	Six fixed cameras	Unsupervised clustering of players Player pose estimation	Four matches	15 f/s

soccer matches a number of doubtful situations arise that cannot be easily judged by the referee committee. Automatic visual systems objectively checking image sequences would prevent incorrect interpretations due to perspective errors, occlusions, or the high velocity of events. In Table 6 we summarize the papers that work on real time event analysis. Broadcast images cannot always be used by these systems because the position of the cameras do not guarantee the exact evaluation of the player position in the playing field. In the last years few research groups have been working on visual frameworks that use fixed cameras with the aim of recognizing real time events, such as ghost goal [86], and offside infringement [87–89].

In [86] the authors present a real time visual system for goal detection which can be used as decision support by the referee committee. Four fixed high resolution cameras acquire 200 f/s that are elaborated by four parallel processes to track the ball and reconstruct the 3D trajectory in order to detect the frame in which the ball passes the goal mouth plane. Objects having features such as circular contour and ball texture are searched for in the image and are considered as ball candidates. A set of neural networks, trained to recognize the ball in different parts of the image and also in different light conditions, are used to select the most probable candidate to contain the ball. A system for automatic judgment of offside events is presented in [87]. The authors propose the usage of 16 cameras located along both sides of the soccer field to cover all the area. Player tracking and team classification are used for offside line calculation; 3D ball trajectory reconstruction allows play recognition; the integration of results from multiple cameras is used for offside judgment. In [88] preliminary results on a small number of images for offside analysis were presented. A method for ball, player and referee detection, team identification and field extraction is proposed. The differentiation between the players of the two teams and the referee is based on the colors of the player uniforms which are compared with three sets of colors manually selected by the user clicking on the jersey of a single player for each team. Projecting the lines from each object onto the global reference point obtained during the field extraction process, the system evaluates the position of all players relative to each other. Six fixed cameras were used in [89] to cover the whole field and to acquire image sequences from both sides of the stadium. Player and ball tracking processes run in parallel on the six image sequences and extract the players and ball position in real time. All the information is projected onto a virtual field by a central supervisor (see Fig. 12). The players' poses extracted from opposite views are related to the ball trajectory in order to detect the player who kicked the ball.

The restricted number of systems for real time event analysis is certainly due to the evident difficulty of processing on-line images acquired with a frame rate that could be, according to the event complexity, higher than standard cameras (30 f/s). Many aspects have to be considered for the application of these systems in real situations, such as their functioning under varying lighting conditions and for the duration of the entire game, high precision in player and ball localization, accuracy in ball–player interaction detection. Just two of these works ware tested on full matches and consider all the aspects related to the on-line processing of 90 min of image sequences.

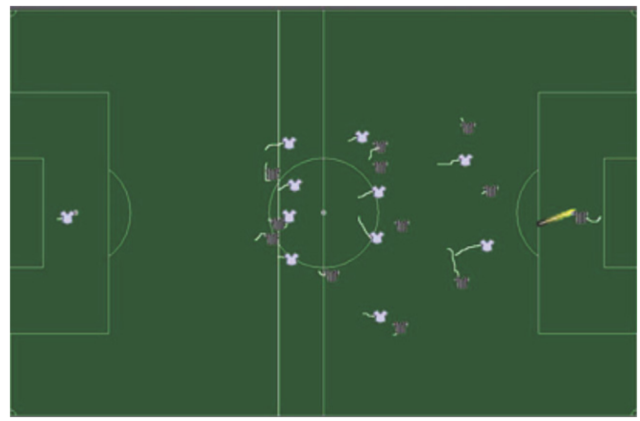

Fig. 12. The exact localization in the virtual field of the interaction between the forward player and the ball during an offside event (from [89]).

5. Discussion

The worldwide popularity of soccer among team sports is certainly due to its simplicity: there is no need of expensive kit and to enjoy watching the game takes very little time as grasping the basic rules is relatively straightforward compared to many other sports. For this reason, soccer video analysis has attracted much research in the last decade. In this paper we present a survey of vision-based system for soccer video analysis. The soccer context is very challenging for the computer vision community since many practical issues must be considered. Depending on the range of application and interest, different methods seem to be more appropriate. In video summarization, the main problem is the extraction of significant events from broadcast images for automatic video clip generation. These systems do not require a semantic analysis of the images because event detection can be done by analyzing the common characteristics of soccer videos during and after significant events, such as the presence of slow motion images, people on the stands celebrating, close up views of referees, and so on. Methods based on the extraction of low-level visual features such as camera motion and dominant color, and on high-level processing to model the feature appearance such as SVM, HMM, CMC have gained a large popularity among these applications. Since they are used for off-line processing, they do not consider the computational cost and often require manual intervention for the initial parameter estimation.

Methods for the provision of augmented information have a large potential market among broadcast television companies since they can enrich the provided services with more information such as statistics, graphical overlays with position information of players, distance covered, velocity and so on, or they can extract customized videos, enriched with semantic information, for live event alerts. In these applications the scene content needs to be interpreted and semantic objects need to be precisely extracted. Many works try to calibrate broadcast images extracting the white lines and feature points to evaluate the camera parameters. The knowledge of the homographic transformation can be used to evaluate the distance between the players, the ball, and the goal. Other works propose the usage of multiple fixed cameras to produce a virtual view synthesis from any point of view or image mosaicing. The common problem of these systems is the a priori knowledge that requires the human intervention and supervision for manual selection of reference points, uniform colors, or the manual input of the gamelog.

The problem of event detection has received much attention due to the increasing computing power in the last decade. The demand for systems capable of detecting high-level events has stimulated a lot of research into the analysis of soccer video. These applications require a semantic analysis of both the players and the ball and their interactions in the field in order to understand the semantics of the considered events. For this reason many works have centered their attention on the ball detection and tracking problem, others on the player segmentation and tracking, and others have considered the problem of putting together the data to discover the interactions.

Trajectory based approaches have been largely used to detect the ball in broadcast images. They evaluate among several candidates the most probable one according to their compatibility with the expected ball trajectory. Generally these methods are off-line and even considering possible future optimizations they are not able to recognize the ball until the whole trajectory is analyzed. Approaches based on pattern recognition techniques cannot be applied unless close up view cameras are used, since the ball texture has to be visible in the images.

Single view player tracking methods are not always able to solve merge group situations. Graph-based approaches consider the relations among players to infer in a backward analysis the possible configurations of players during occlusions. Model-based approaches use colors to distinguish players of different teams, however, they suffer when they have to distinguish among occluding players of the same team. These problems are for the most part solved by multicamera systems that fuse information from single view analysis. Player labeling approaches generally require the manual selection of some initial patches needed for the model extraction phase, or the a priori knowledge of significant features that could be used for team recognition during tests.

A few works have been presented in literature for game analysis and team statistics. The proposed systems are not completely automatic but they can greatly improve the manual work during post processing analysis. Some of them require human intervention for the initial selection of players, or for disambiguating complex situations such as player occlusions. Tests are often done on simulation sequences and no mention is made of computational costs.

Recently, a few papers have been published for real time event detection. Automatic systems that analyze the players position in the field during matches can be used by referees as support for their decisions. These systems generally use fixed cameras since broadcast images do not guarantee the appropriate positioning for the solution of perspective problems. Methodologies for ball detection, player tracking and classification, have been integrated with homographic transformation procedures to extract the exact positions in the virtual field and detect the ball–player interactions. In order to be applicable to real time processing these methodologies are a tradeoff between performance and computational cost. Two works have been presented that follow this research direction.

5.1. Challenges ahead

An important issue in the realization of successful soccer video analysis systems is the design of algorithms that yield a maximum level of reliability and robustness. Although many papers introduced visual systems for video analysis in several application contexts, in our opinion a great deal of work should be directed towards the enhancement of automatic analysis to reduce manual intervention and improve their performance. Even if many methods are quite effective on some test sequences, highly robust reliable systems are yet to be demonstrated as being appropriate for extensive tests and for long term applications. Soccer matches can be played at different times of the day with natural or artificial light and also during the same match the lighting conditions can vary greatly, changing suddenly from sunny to cloudy to rainy and so on. Automatic systems have to be robust to manage different and evolving light conditions.

Achieving these objectives requires the addressing of several challenges and solving many different problems. Developing more robust algorithms for ball and player detection is the first essential issue. These algorithms have to operate robustly and reliably in complex and widely varying situations (sunny, cloudy, rainy and so on). Tracking algorithms have to be able to manage complex group situations, with players occluding each other and the ball being out of the field of view for long periods. Also for the automatic construction of player models we believe that several orders of improvement are needed to improve the performances of algorithms and to guarantee the effectiveness in any configuration of team uniforms. The above issues are fundamental to the successful application of any further analysis such as player interactions, tactic analysis, real time event detection and so on. Determining the desired level of accuracy for these steps is not easy and depends on the nature of the application. In Fig. 13 we report the graph that relates the processing time with the ball player detection accuracy that should be reached in different application contexts. On the x axis the processing time was divided into three possible ranges off-line, on-line (the processing results have to be provided with few seconds delay), and real time (the processing results have to be provided within tenths of a second). On the y axis the accuracy was divided into three ranges: low, medium (the relative positions of players and the covered distances are important) and high (the precision of the ball and player localizations in the field is crucial). Video summarization applications are generally off-line and do not require a great accuracy in the ball player detection since the highlight selection can be carried out extracting low-level features such as color, camera motion, and audio features. Applications for the provision of augmented information have to be on line in order to provide useful services to broadcast television companies or mobile phone companies. In these cases, the ball and player detection algorithms need to be highly accurate and have to provide results with only a few seconds delay. Also team statistical analysis applications could provide valuable suggestions to team coaches and trainers if on-line results were produced. On the contrary, systems for real time events applications are requires to be as accurate as possible both in ball and in players localization. In order to support referees during matches, automatic systems for event interpretation such as goal detection, automatic

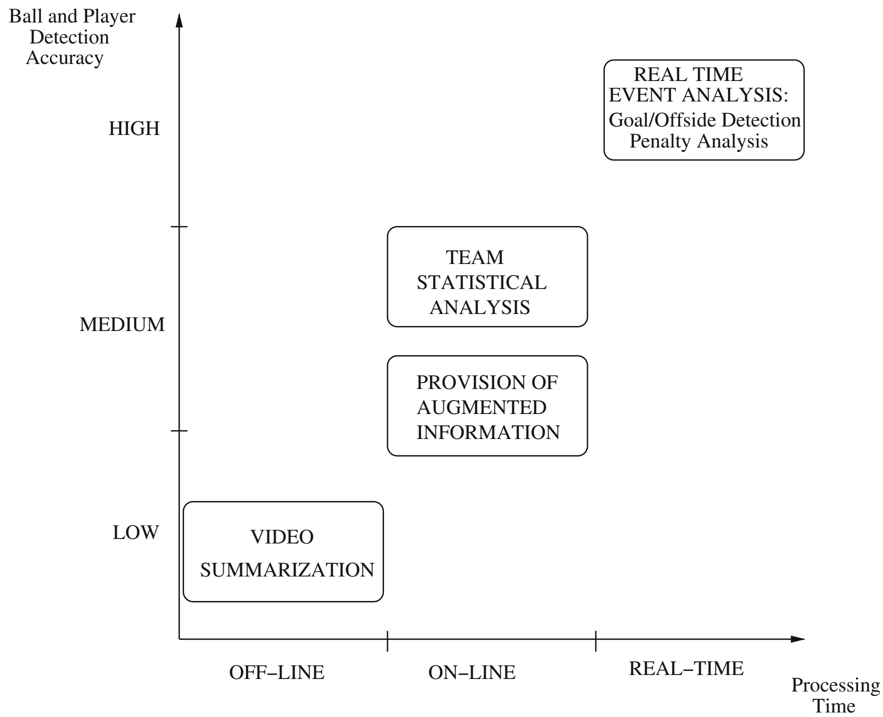

Fig. 13. The expected collocation of different soccer applications with respect to the processing time and the localization accuracy.

judgement of offside actions, have to detect with high accuracy the ball position with respect the goal line, the distance between players in the same frames in which shots were recognized, and the team the players belongs to. These applications are certainly the most difficult from a technical point of view, since their successful utilization will depend on the number of correct ball/player detections and on the processing time. In our opinion, while some algorithms (such as background subtraction, camera calibration, feature extraction) can be easily optimized to meet real time requirements, there are other methodologies that cannot be considered for on-line and real time applications (such as ball detection by trajectory evaluation, or player tracking by searching paths in graph hierarchies). In these cases a trade off between the methodology's complexity and the processing constraints has to be considered.

From our perspective, the future holds promise for visual analysis systems. Soccer is one of the most popular sports in the world. It receives a great deal of interest both from the audience during matches, and because of the large quantity of television programs dedicated to the diffusion of events and debates on the results. Major mobile telecommunication companies compete to offer new customized services to respond to the increasing requirements of prospective clients. We believe that significant progress is going to be made over the next few years: the experience of the computer vision community in similar contexts combined with the increasing sophistication of computational and sensing devices can produce in the future more reliable systems for the continued growth of this research field.

References

[1] R. Leonardi, P. Migliorati, M. Prandini, Semantic indexing of soccer audio-visual sequences: a multimodal approach based on controlled Markov chains, IEEE Transactions on Circuits and Systems for Video Technology 14 (5) (2004) 634–643.
[2] Y.-L. Kang, J.-H. Lim, M.S. Kankanhalli, C.-S. Xu, Q. Tian, Goal detection in soccer video using audio/visual keywords, in: IEEE International Conference on Image Processing (ICIP), Singapore, 24–27 October 2004, pp. 1629–1632.
[3] D.A. Sadlier, O. Connor, Event detection in field sports video using audio-visual features and a support vector machine, IEEE Transactions on Circuits and Systems for Video Technology 5 (10) (2005) 1225–1233.
[4] S.C. Chen, M.L. Shyu, M. Chen, C. Zhang, A decision tree-based multimodal data mining framework for soccer goal detection, in: IEEE Proceedings of International Conference on Multimedia and Expo (ICME), Taipei, Taiwan, 27–30 June 2004, pp. 265–268.
[5] L. Xie, P. Xu, S.F. Chang, A. Divakaran, H. Sun, Structure analysis of soccer video with domain knowledge and hidden Markov models, Pattern Recognition Letters 25 (7) (2004) 767–775.
[6] A. Ekin, A. Tekalp, R. Mehrotra, Automatic soccer video analysis and summarization, IEEE Transactions on Image Processing 12 (2003) 796–807.
[7] Z. Aghbari, K. Kaneko, A. Makinouchi, Content-trajectory approach for searching video databases, IEEE Transactions on Multimedia 5 (4) (2003) 516–531.
[8] F. Wang, Y.F. Ma, H.J. Zhang, J.T. Li, A generic framework for semantic sports video analysis using dynamic Bayesian networks, in: Proceedings of the 11th International Multimedia Modelling Conference, Melbourne, Australia, 12–14 January 2005, pp. 115–122.
[9] C.L. Huang, H.C. Shih, C.Y. Chao, Semantic analysis of soccer video using dynamic Bayesian network, IEEE Transactions on Multimedia 8 (4) (2006) 749–760.
[10] Y. Gao, W.B. Wang, J.H. Yong, H.J. Gu, Dynamic video summarization using two-level redundancy detection, Multimedia Tools and Applications 42 (2009) 233–250.
[11] J. Assfalg, M. Bertini, C. Colombo, A. Del Bimbo, W. Nunziati, Semantic annotation of soccer videos: automatic highlights identification, Computer Vision and Image Understanding 92 (2–3) (2003) 285–305.
[12] Y. Ariki, S. Kubota, M. Kumano, Automatic production system of soccer sports video by digital camera work based on situation recognition, in: Proceedings of the Eight IEEE International Symposium on Multimedia (ISM'06), San Diego, California, 11–13 December 2006.
[13] C. Xu, Y.F. Zhang, G. Zhu, Y. Rui, H. Lu, Q. Huang, Using webcast text for semantic event detection in broadcast sports video, IEEE Transaction on Multimedia 10 (7) (2008) 1342–1355.
[14] R. Gonzales, R. Woods, Digital Image Processing, third ed., Prentice-Hall, Upper Saddle River, NJ, 2008.
[15] I. Otsuka, K. Nakane, A. Divakaran, K. Hatanaka, M. Ogawa, A highlight scene detection and video summarization system using audio feature for a personal video recorder, IEEE Transactions on Consumer Electronics 51 (1) (2005) 112–116.
[16] C.C. Cheng, C.T. Hsu, Fusion of audio and motion information on HMM-based highlight extraction for baseball games, IEEE Transactions on Multimedia 8 (3) (2006) 585–599.
[17] D.W. Tjondronegoro, Y.-P. Chen, B. Pham, Classification of self-consumable highlights for soccer video summaries, in: IEEE International Conference on Multimedia and Expo, Baltimore, Maryland, 6–9 July 2003, pp. 579–582.
[18] Y.Q. Yang, Y.D. Lu, W. Chen, A framework for automatic detection of soccer goal event based on cinematic template, in: Proceedings of 2004

International Conference on Machine Learning and Cybernetics, Shanghai, China, 26–29 August 2004, pp. 3759–3764.
[19] K.W. Seo, J.S. Ko, I.K. Ahn, C.G. Kim, An intelligent display scheme of soccer video on mobile devices, IEEE Transactions on Circuits and Systems for Video Technology 17 (10) (2007) 1395–1401.
[20] C. Pei, L. Gao, S. Yang, C. Hou, A ROI detection model for soccer video on small display, in: Third International Symposium on Intelligent Information Technology Application, vol. 2, Nanchang University, China, 21–22 November 2009, pp. 392–395.
[21] M. Bertini, R. Cucchiara, A. Bimbo, A. Prati, Semantic adaptation of sport videos with user-centred performance analysis, IEEE Transactions on Multimedia 8 (3) (2006) 433–443.
[22] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, Cambridge, UK, 2000.
[23] O. Faugeras, Three-Dimensional Computer Vision: A Geometric Approach, MIT Press, Cambridge, MA, 1996.
[24] R.Y. Tsai, A versatile camera calibration technique for 3D machine vision, IEEE Journal Robotics & Automation (4) (1987) 323–344.
[25] A. Yilmaz, O. Javed, M. Shah, Object tracking: a survey, ACM Computing Surveys 38 (4) (2006) 13–58.
[26] D. Ramanan, D.A. Forsyth, A. Zisserman, Tracking people by learning their appearance, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (1) (2007) 65–81.
[27] E.L. Andrade, J.C. Woods, E. Khan, M. Ghanbari, Region-based analysis and retrieval for tracking of semantic objects and provision of augmented information in interactive sport scenes, IEEE Transactions on Multimedia 7 (6) (2005) 1084–1096.
[28] G. Thomas, Real-time camera tracking using sports pitch markings, Journal Real Time Image Processing 2 (2–3) (2007) 117–132.
[29] X. Yu, X. Yan, L. Li, H.W. Leong, An instant semantics acquisition system of live soccer video with application to live event alert and on-the-fly language selection, in: Conference on Image and Video Retrieval (CIVR), Niagara Falls, Canada, 7–9 July 2008.
[30] T. Rodriguez, I.D. Reid, R. Horaud, N. Dalal, M. Goetz, Image interpolation for virtual sports scenarios, Machine Vision and Applications 16 (4) (2005) 236–245.
[31] N. Inamoto, H. Saito, Virtual viewpoint replay for a soccer match by view interpolation from multiple cameras, IEEE Transactions on Multimedia 9 (6) (2007) 1155–1166.
[32] J.B. Hayet, J. Piater, J. Verly, Robust incremental rectification of sports video sequences, in: British Machine Vision Conference 2004, Kingston University, London, 7–9 September 2004.
[33] J. Wang, C. Xu, E. Chng, H. Lu, Q. Tian, Automatic composition of broadcast sports video, Multimedia Systems 14 (4) (2008) 179–193.
[34] C.R. Jung, Efficient background subtraction and shadow removal for monochromatic video sequences, IEEE Transactions on Multimedia 11 (3) (2009) 571–577.
[35] J.M. McHugh, J. Konrad, V. Saligrama, P.M. Jodoin, Foreground-adaptive background subtraction, IEEE Signal Processing Letters 16 (5) (2009) 390–393.
[36] J. Zhang, S. Gong, People detection in low-resolution video with non-stationary background, Image and Vision Computing 27 (4) (2009) 437–443.
[37] S.M. Khan, M. Shah, Tracking multiple occluding people by localizing on multiple scene planes, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (3) (2009) 505–519.
[38] F. Fleuret, J. Berclaz, R. Lengagne, P. Fua, Multicamera people tracking with a probabilistic occupancy map, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (2) (2008) 267–282.
[39] J. Black, T. Ellis, Multi camera image tracking, Image and Vision Computing 24 (11) (2006) 1256–1267.
[40] C. Piciarelli, G.L. Foresti, On-line trajectory clustering for anomalous events detection, Pattern Recognition Letters 27 (15) (2006) 1835–1842.
[41] I. Haritaoglu, D. Harwood, L. Davis, W4: real-time surveillance of people and their activities, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8) (2000) 809–830.
[42] C. Stauffer, W. Grimson, Adaptive background mixture models for real-time tracking, in: International Conference on Compute Vision and Pattern Recognition CVPR, Ft. Collins, CO, USA, June 1999, pp. 246–252.
[43] M. Xu, T. Ellis, Partial observation vs. blind tracking through occlusion, in: Proceedings of the British Machine Vision Conference, Cardiff 2–5 September 2002, pp. 777–786.
[44] J. Ren, J. Orwell, G. Jones, M. Xu, Real-time modeling of 3-d soccer ball trajectories from multiple fixed cameras, IEEE Transaction on Circuits and Systems for Video Technology 18 (3) (2008) 350–362.
[45] X. Yu, H.W. Leong, C. Xu, Q. Tian, Trajectory-based ball detection and tracking in broadcast soccer video, IEEE Transactions on Multimedia 8 (6) (2006) 1164–1178.
[46] Y. Liu, D. Liang, Q. Huang, W. Gao, Extracting 3D information from broadcast soccer video, Image and Vision Computing 24 (10) (2006) 1146–1162.
[47] T. Shimawaki, T. Sakiyama, J. Miura, Y. Shirai, Estimation of ball route under overlapping with players and lines in soccer video image sequence, in: International Conference on Pattern Recognition ICPR, Hong Kong, 20–24 August 2006, pp. 359–362.
[48] K. Choi, Y. Seo, Tracking soccer ball in TV broadcast video, Image Analysis and Processing (ICIAP), Cagliari, Italy, 6–8 September 2005, pp. 661–668.
[49] V. Pallavi, J. Mukherjee, A.K. Majumdar, S. Sural, Ball detection from broadcast soccer videos using static and dynamic features, Journal Visual Communication and Image Representation 19 (7) (2008) 426–436.
[50] T. D'Orazio, M. Leo, A. Distante, C. Guaragnella, New algorithm for ball recognition using circle hough transform and neural classifier, Pattern Recognition 37 (3) (2004) 393–408.
[51] T. Misu, A. Matsui, M. Naemura, M. Fujii, N. Yagi, Distributed particle filtering for multiocular soccer ball tracking, in: IEEE International Conference on Acoustic, Speech and Signal Processing, Hawaii, USA, 15–20 April 2007, pp. 937–940.
[52] P.J. Figueroa, N.J. Leite, L.R.M. Barros, Background recovering in outdoor image sequences: an example of soccer player segmentation, Image and Vision Computing 24 (4) (2006) 363–374.
[53] J.R. Renno, J. Orwell, D.J. Thirde, G.A. Jones, Shadow classification and evaluation for soccer player detection, in: British Machine Vision Conference, Kingston University, London, 7–9 September 2004, pp. 839–848.
[54] N. Vandenbroucke, L. Macaire, J.G. Postaire, Color image segmentation by pixel classification in an adapted hybrid color space. Application to soccer image analysis, Computer Vision and Image Understanding 90 (2) (2003) 190–216.
[55] S.H. Khatoonabadi, M. Rahmati, Automatic soccer players tracking in goal scenes by camera motion elimination, Image and Vision Computing 27 (4) (2009) 469–479.
[56] M. Naemura, A. Fukuda, Y. Mizutani, Y. Izumi, Y. Tanaka, K. Enami, Morphological segmentation of sport scenes using color information, IEEE Transactions on Broadcasting 46 (3) (2000) 181–188.
[57] M. Beetz, N.v. Hoyningen-Huene, J. Bandouch, B. Kirchlechner, S. Gedikli, A. Maldonado, Camera-based observation of football games for analyzing multi-agent activities, in: AAMAS '06: Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems, Hakodate, Japan, 8–12 May 2006, pp. 42–49.
[58] S. Gedikli, J. Bandouch, N. Hoyningen-Huene, B. Kirchlechner, M. Beetz, An adaptive vision system for tracking soccer players from variable camera settings, in: Proceedings of the Fifth International Conference on Computer Vision Systems (ICVS 2007), Bielefeld University, Germany, 21–24 March 2007.
[59] M. Beetz, S. Gedikli, J. Bandouch, B. Kirchlechner, N. von Hoyningen-Huene, A. Perzylo, Visually tracking football games based on TV broadcasts, in: International Joint Conference on Artificial Intelligence, Hyderabad, India, 6–12 January 2007, pp. 2066–2071.
[60] J. Liu, X. Tong, W. Li, T. Wang, Y. Zhang, H. Wang, B. Yang, L. Sun, S. Yang, Automatic player detection, Labeling and Tracking in Broadcast Soccer Video, in: British Machine Vision Conference, University of Warwick, UK, 10–13 September 2007.
[61] J. Liu, X. Tong, W. Li, T. Wang, Y. Zhang, H. Wang, Automatic player detection, labeling and tracking in broadcast soccer video, Pattern Recognition Letters 30 (2) (2009) 103–113.
[62] P. Nillius, J. Sullivan, S. Carlsson, Multi target tracking—linking identities using Bayesian network inference, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), New York, USA, 17–22 June 2006.
[63] J. Sullivan, S. Carlsson, Tracking and labelling of interacting multiple targets, in: Proceedings of Ninth European Conference on Computer Vision (ECCV), Graz, Austria, 7–13 May 2006.
[64] S.W. Joo, R. Chellappa, A multiple hypothesis approach for multiobject tracking, IEEE Transaction on Image Processing 16 (11) (2007) 2849–2854.
[65] J. Miura, H. Kubo, Tracking players in highly complex scenes in broadcast soccer video using a constraint satisfaction approach, in: Proceedings of CIVR 2008, Niagara Falls, Canada, 2008.
[66] V. Pallavi, J. Mukherjee, A.K. Majumdar, S. Sural, Graph based multiplayer detection and tracking in broadcast soccer video, IEEE Transactions on Multimedia 10 (2008) 794–805.
[67] M. Xu, J. Orwell, L. Lowey, D. Thirde, Architecture and algorithms for tracking football players with multiple cameras, IEE Proceedings—Vision, Image and Signal Processing 152 (2) (2005) 232–241.
[68] W. Du, J.B. Hayet, J. Piater, J. Verly, Collaborative Multicamera tracking of athletes in Team Sports, CVBASE '06—Workshop on Computer Vision Based Analysis in Sport Environments, Graz, Austria, 12 May 2006.
[69] W. Du, J. Piater, Multi-camera people tracking by collaborative particle filters and principal axis-based integration, in: Lecture Notes in Computer Science, vol. 4843, 2007, pp. 365–374.
[70] P.J. Figueroa, N.J. Leite, L.R.M. Barros, Tracking soccer players aiming their kinematical motion analysis, Computer Vision and Image Understanding 101 (2) (2006) 122–135.
[71] T. Misu, S. Gohshi, Y. Izumi, Y. Fujita, M. Naemura, Robust tracking of athletes using multiple features of multiple views, in: International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, Plzen-Bory, Czech Republic, 2–6 February 2004, pp. 285–292.
[72] J.B. Hayet, T. Mathes, J. Czyz, J. Piater, J. Verly, B. Macq, A modular multi-camera framework for team sports tracking, in: IEEE Conference on Advanced Video and Signal Based Surveillance, Como, Italy, 15–16 September 2005, pp. 493–498.
[73] T. Mathes, J. Piater, Robust non-rigid object tracking using point distribution models, British Machine Vision Conference, Oxford, UK, 5–8 September 2005.
[74] P. Spagnolo, P.L. Mazzeo, M. Leo, T. D'Orazio, Unsupervised algorithms for segmentation and clustering applied to soccer pla rs classification, in: Proceeding of the International Conference on Signal Processing and Multimedia Applications, Barcelona, Spain, 28–31 July 2007, pp. 129–134.

[75] H. Kim, S. Nam, J. Kim, Player segmentation evaluation for trajectory estimation in soccer games, in: Proceeding of Image and Vision Computing, Palmerston North, New Zealand, 26–28 November 2003, pp. 159–162.
[76] M. Kristan, J. Pers, M. Perse, S. Kovacic, M. Bon, Multiple interacting targets tracking with application to team sports, in: Proceeding of the Fourth International Symposium on Image and Signal Processing and Analysis, Zagreb, Croatia, 15–17 September 2005.
[77] Z. Xu, P. Shi, Segmentation of player and team discrimination in soccer video, in: Proceedings of the IEEE International Workshop on VLSI Design and Video Technology (IWVDVT), Suzhou, China, 28–31 May 2005.
[78] S. Iwase, H. Saito, Parallel tracking of all soccer players by integrating detected positions in multiple view images, in: ICPR '04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04), vol. 4, Cambridge, UK, 23–26 August 2004, pp. 751–754.
[79] G. Zhu, Q. Huang, C. Xu, Y. Rui, S. Jiang, W. Gao, H. Yao, Trajectory based event tactics analysis in broadcast sports video, in: MULTIMEDIA '07: Proceedings of the 15th International Conference on Multimedia, Augsburg, Germany, 24–29 September 2007, pp. 58–67.
[80] G. Zhu, C. Xu, Y. Zhang, Q. Huang, H. Lu, Event tactic analysis based on player and ball trajectory in broadcast video, in: Conference on Image and Video Retrieval (CIVR), Niagara Falls, Canada, 7–9 July 2008.
[81] R.M.L. Barros, M.S. Misuta, R.P. Menezes, P.J. Figueroa, F.A. Moura, S.A. Cunha, R. Anido, N.J. Leite, Analysis of the distances covered by first division Brazilian soccer players obtained with an automatic tracking method, Journal of Sports Science and Medicine 6 (2) (2007) 233–242.
[82] X. Yu, T. Sen Hay, X. Yan, E. Chng, A player possession acquisition system for broadcast soccer video, in: Proceedings of the IEEE International Conference on Multimedia & Expo, Amsterdam, The Netherlands, 6–8 July 2005.
[83] P.S. Tsai, T. Meijome, P.G. Austin, Scout: a game speed analysis and tracking system, Machine Vision and Application 8 (5) (2007) 289–299.
[84] C. Kang, J. Hwang, N.K. Li, Trajectory analysis for soccer players, in: Proceedings of the Sixth IEEE International Conference on Data Mining—Workshops (ICDMW'06), Hong Kong, China, 18 December 2006.
[85] M. Beetz, B. Kirchlechner, M. Lames, Computerized real-time analysis of football games, IEEE Pervasive Computing 4 (3) (2005) 33–39.
[86] T. D'Orazio, M. Leo, P. Spagnolo, M. Nitti, N. Mosca, A. Distante, A visual system for real time detection of goal events during soccer matches, Computer Vision and Image Understanding 113 (5) (2009) 622–632.
[87] S. Hashimoto, S. Ozawa, A system for automatic judgment of offsides in soccer games, in: IEEE International Conference on Multimedia and Expo, Toronto, Canada, 9–12 July 2006, pp. 1889–1892.
[88] W.C. Naidoo, J.R. Tapamo, Soccer video analysis by ball, player and referee tracking, in: SAICSIT '06: Proceedings of the 2006 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries, Somerset West, South Africa, 9–11 October 2006, pp. 51–60.
[89] T. D'Orazio, M. Leo, P. Spagnolo, P.L. Mazzeo, N. Mosca, M. Nitti, A. Distante, An investigation into the feasibility of real-time soccer offside detection from a multiple camera system, IEEE Transaction on Circuits and Systems for Video Technology 19 (12) (2009) 1804–1817.

About the Author—TIZIANA D'ORAZIO received the Laurea degree magna cum laude in 1988 in Computer Science from the University of Bari. She held grants from the Italian National Council (CNR) of Bari in 1989–1993 for research activities in Robotics and Image Processing. Since 1997 she has been a Researcher at the Institute of Intelligent Systems for Automation (ISSIA) of the Italian National Research Council. She participated to some CNR and MIUR projects and was responsible for some collaborations between the ISSIA Institute and industrial companies. Her current research interests include pattern recognition, video analysis and computer vision for video surveillance, domotics, Intelligent Transportation Systems, and Quality Control. Past activity is concerned with visual algorithms for autonomous navigation, artificial intelligence paradigm for elementary behavior learning, feature extraction for pattern recognition. Video surveillance activity is devoted to indoor and outdoor application for people and posture tracking and behavior understanding. She is responsible of many projects with industrial partners and private companies in pattern recognition, computer vision, and visual surveillance. She is author of more than 100 papers in national and international journals, and conference proceedings. She is Co-author of three international patents on visual systems for event detection in sport contexts.

About the Author—MARCO LEO received an Honors degree in Computer Science Engineering from the University of Lecce in 2001. Since then, He is a Researcher at the Italian National Research Council (C.N.R.), Institute of Intelligent Systems for Automation (ISSIA) in Bari (Italy). His main research interests are in the fields of image processing, image analysis, computer vision, pattern recognition, digital signal processing, neural networks, graphical models, wavelet transform and independent component analysis. He participated to a number of national and international research projects facing automatic video surveillance of indoor and outdoor environments, human attention monitoring, real time event detection in sport contexts and non-destructive inspection of aircraft components. He is author of more than 100 papers in national and international journals, and conference proceedings. He is Co-author of national and international patents on visual systems for event detection in sport contexts.