

Ball tracking in sports: a survey

Paresh R. Kamble¹ · Avinash G. Keskar¹ ·
Kishor M. Bhurchandi¹

Published online: 16 October 2017
© Springer Science+Business Media B.V. 2017

Abstract Increase in the number of sport lovers in games like football, cricket, etc. has created a need for digging, analyzing and presenting more and more multidimensional information to them. Different classes of people require different kinds of information and this expands the space and scale of the required information. Tracking of ball movement is of utmost importance for extracting any information from the ball based sports video sequences. Based on the literature survey, we have initially proposed a block diagram depicting different steps and flow of a general tracking process. The paper further follows the same flow throughout. Detection is the first step of tracking. Dynamic and unpredictable nature of ball appearance, movement and continuously changing background make the detection and tracking processes challenging. Due to these challenges, many researchers have been attracted to this problem and have produced good results under specific conditions. However, generalization of the published work and algorithms to different sports is a distant dream. This paper is an effort to present an exhaustive survey of all the published research works on ball tracking in a categorical manner. The work also reviews the used techniques, their performance, advantages, limitations and their suitability for a particular sport. Finally, we present discussions on the published work so far and our views and opinions followed by a modified block diagram of the tracking process. The paper concludes with the final observations and suggestions on scope of future work.

Keywords Tracking set ups · Ball detection · Shape and size features · Tracking performance evaluation · Ball tracking · Tracking techniques classification

✉ Paresh R. Kamble
paresh.kamble@students.vnit.ac.in; kamble.paresh@gmail.com

¹ Department of Electronics and Communication Engineering, Visvesvaraya National Institute of Technology, Nagpur, India

1 Introduction

Huge efforts have been put in tracking, understanding and visualization of events from sports environments. Due to increase in the number and variety of sport lovers and their interest, there is a great demand of obtaining multidimensional data from the game. Public may need to view the events from arbitrary viewpoint. Amateurs need to understand the precise movements of the professional sportsmen to tune their own plans or moves. Sports persons and coaches need to analyze the tactical movements of the opposition players and teams in order to design their own counter plan. Broadcasting and publishing agencies need to detect, annotate, store, categorize and summarize the available video sequences based on the events occurring during the game. This should facilitate easy retrieval of a particular match event or moment within no time and produce highlights as and when required. Since ball is the center of attention and attraction of the games considered in this work, its precise tracking helps facilitate the above said tasks and hence we present a comprehensive survey of tracking of a ball.

We categorized ball games with respect to similarities in challenges including different level of difficulties due to size, randomness, speed of movement of the ball, etc. and shown in Table 1.

Thus, the task of tracking the sport entities is not problem-free. There are a large number of challenges and problems involved which include occlusion, misdetection, illumination variation, different quality of videos with varying frame rates, real-time constraint, etc.

Table 1 Ball games categorized as per similarities in challenges and issues

Sport	Dominant features/advantages	Challenges
Pool	<ul style="list-style-type: none"> • Small region needed 	<ul style="list-style-type: none"> • Small ball size
Billiards	<ul style="list-style-type: none"> • Ball color 	<ul style="list-style-type: none"> • Varying ball velocity
Table Tennis	<ul style="list-style-type: none"> • Less cameras needed 	<ul style="list-style-type: none"> • Non-linear, non-periodic oscillatory motion
Tennis	<ul style="list-style-type: none"> • Court lines (Tennis) 	<ul style="list-style-type: none"> • Medium ball size
Golf	<ul style="list-style-type: none"> • Little or no occlusion • Less cameras needed 	<ul style="list-style-type: none"> • High ball velocity
Volleyball	<ul style="list-style-type: none"> • Bigger ball size • Slow ball velocity 	<ul style="list-style-type: none"> • Same team occlusion
Basketball	<ul style="list-style-type: none"> • Bigger ball size 	<ul style="list-style-type: none"> • Same and opposite team occlusion
Football	<ul style="list-style-type: none"> • Lines and circles on ground 	<ul style="list-style-type: none"> • Player can possess the ball for long time • More cameras needed
Hockey	<ul style="list-style-type: none"> • Lines on ground (Baseball) 	<ul style="list-style-type: none"> • Medium ball size
Cricket		<ul style="list-style-type: none"> • Occlusion
Baseball		<ul style="list-style-type: none"> • High ball velocity • More cameras needed
American Football	<ul style="list-style-type: none"> • Lines on ground 	<ul style="list-style-type: none"> • Very heavy occlusion
Rugby		<ul style="list-style-type: none"> • Spheroidal ball shape • Irregular trajectories

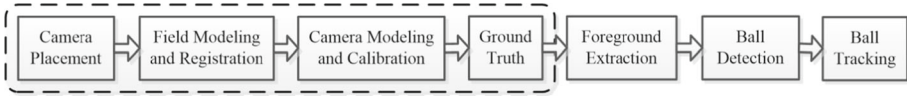


Fig. 1 Overview of a ball tracking system

They produce different levels of difficulties for the ball tracking algorithms. This paper presents a comprehensive survey of various detection and tracking techniques for a ball in various sport events. Wherever possible the quantitative performances of the published algorithms have been presented using well-known measures like accuracy, precision, recall and F-score. This survey highlights the physical infrastructure, algorithmic approaches and steps, various milestones in tracking techniques, research and performance benchmarking of the available ball tracking algorithms. We also propose a modified model of a ball tracking system.

The ball tracking system is composed of large number of stages that include system modeling and pre-processing blocks, foreground extraction, followed by ball detection and ball tracking. Figure 1 shows the conventional block diagram of a ball tracking system.

The conventional system invariably requires certain pre-processing in order to get ready for actual detection and tracking. Region within dashed line in Fig. 1 shows the pre-processing stage. Its first stage consists of placement of camera(s) and acquisition of their initialization parameters. Stage 2 constructs a model of the game field obtained from the dimensions provided by rulebook of the game. The third stage involves camera modelling and calibration of each camera with homographic calculations for registering the play field that involves extracting the field lines and the feature points. Stage 4 further constructs a well-defined set of ground truths using the game rules.

Section 2 presents pre-processing stage with physical setups introduced so far for different ball tracking systems including camera placements, field modeling and registration, processes for camera calibration, ground truth and foreground extraction. Section 3 presents different features and methods used for ball detection techniques introduced so far. Section 4 presents different tracking algorithms under different categories like; Kalman Filter, Particle Filter, Trajectory based, and other miscellaneous types. Section 5 presents discussion on the major milestones in the literature survey, our observations and insight of the developments so far. Section 6 presents concluding remark and future scope.

2 Pre-processing and foreground extraction

In order to get ready for detection and tracking, it is important to set up the system, calibrate it, make it robust to noise and clutter, obtain ground truth and extract foreground. Pre-processing stage serves the purpose. Subsequent text of this section describes each sub-block of the pre-processing module in brief.

2.1 Camera placement

In order to capture the precise movements of the players, bat, racket, ball, etc. on the field or court it is important to employ appropriate number of cameras around it. Though the algorithms are expected to track a ball, other said objects are also required to be tracked in

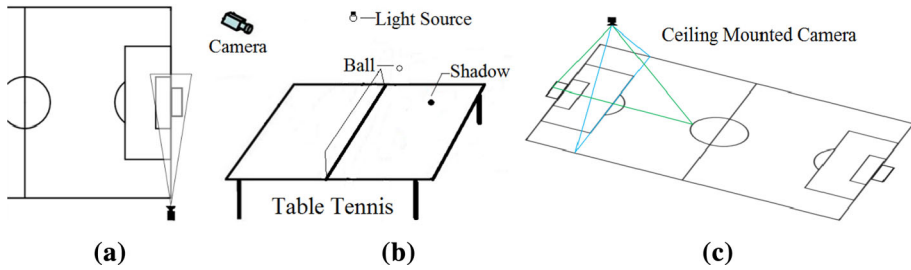


Fig. 2 Different types of single camera systems. **a** Camera's optical axis lying on goal mouth plane (D'Orazio et al. 2004). **b** Camera with light source (Zhang et al. 2011). **c** Single ceiling mounted camera (Santiago et al. 2011)

the dynamically changing area of interest for detailed analysis of events and indemnification of game rules. The video acquisition method has to be non-invasive to keep the rules of the games unchanged and still acquire all the minutest relevant events.

2.1.1 Single camera system

Researchers generally use a single fixed camera for capturing many sports events. The camera is placed on one side of the soccer field with its optical axis lying on the goal-mouth plane (D'Orazio et al. 2004) as shown in Fig. 2a or focused on the goal area (Ancona et al. 2003). Sometimes a fixed camera with properties like no zooming or panning capability (Bloom and Bradley 2003; Chen et al. 2012; Shah et al. 2007) are used. Camera with properties like monochrome (Chen and Wang 2007), low-resolution (Teachabarikiti et al. 2010), high speed (Glover and Kaelbling 2014; Shum and Komura 2005), high vision (Ariki et al. 2008), with light source (Zhang et al. 2011) (Fig. 2b), etc. are used to capture a fixed view of the field. Generally, high-speed cameras track the translational, rotational and spin movements of the ball. In order to reduce the burden of building a two camera based binocular vision system, ceiling mounted cameras (Fig. 2c) are preferred (Perše et al. 2005; Santiago et al. 2011). The single moving camera involves the using Pan-Tilt-Zoom (PTZ) capabilities (Dearden et al. 2006; Han et al. 2003; Yan et al. 2006b). The camera may be implanted on a robotic hand and use it in eye-in-hand configuration for tracking and catching a ball (Cigliano et al. 2015). The camera of a hand-held device (Uchiyama and Saito 2007) or a camcorder (Chakraborty and Meher 2011) may be used to take videos of the sports for tracking purpose.

2.1.2 Two camera system

A general purpose of using a two-camera system is 3D localization of the ball and players. Reducing the effect of occlusion and having clearer views from multiple angles are other major reasons for using the system. It includes using stereo cameras (Birbach and Frese 2009; Zhang et al. 2010) which may be unsynchronized (Liu et al. 2014) (as seen in Fig. 3a) or synchronized (Ishii et al. 2007; Kasuya et al. 2010) (as seen in Fig. 3b, c) or even dynamically coordinated (Hoyningen-Huene and Beetz 2009). The cameras may be placed outside the field at either the average height of player or a comparatively greater height to reduce occlusion (Kasuya et al. 2010) and have a bird's eye view of the court (Misu et al. 2007).

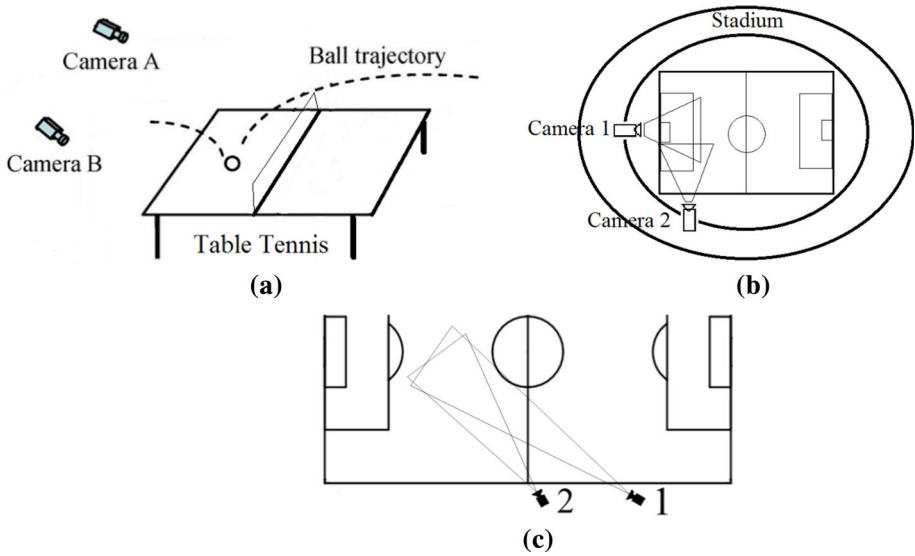


Fig. 3 Two-camera system. **a** Unsyncronized two-camera system in Table Tennis (Liu et al. 2014). **b, c** Two cameras placed at higher position either to have bird's eyeview of field or occlusion reduction (Kasuya et al. 2010; Misu et al. 2007)

2.1.3 Multi-camera systems

Apart from 3D localization of ball, such types of systems are usually employed where the ground is too big to be captured by one or two camera systems. Other purpose of such type of systems can be either to obtain a continuous shot of the complete game field or to cover every nuke and corner of the stadium clearly without occlusion or intrusion. These types of systems employ four cameras (Figueroa et al. 2004) which either may be static (Choi and Seo 2011) or dynamic high speed (D'Orazio et al. 2009a,b) (as seen in Fig. 4a) to cover the whole field. Two out of four high-speed cameras placed on both sides of one goal post are used to observe the goal events. The six-camera system consists of four cameras placed on the sides and two at the court ends (Pingali et al. 2000) at appropriate heights. Every side camera forms a pair with its adjacent end camera as seen in Fig. 4b. It could be part of an 8-monochrome camera system (Pingali et al. 2002) where other two cameras are used for player tracking. Each of the four cameras was paired with the two end cameras in order to observe the ball in 3D. The placement of the six cameras may contain three high-speed moving cameras on each side of the ground (D'Orazio et al. 2009a,b; Leo et al. 2008) so that all the area is doubly covered either by the adjacent or the opposite cameras. ISSIA dataset, prepared in accordance with this arrangement as in Leo et al. (2008), was used in Wang et al. (2014a) as in Fig. 4c. The APIDIS dataset of basketball match was prepared using 7 unsyncronized cameras placed one at the top and six around the court.

To obtain the multi-camera shots, Iwase and Saito (2003) used eight stable cameras with four lined up and set at both sides of the ground and aiming at the penalty area, as shown in Fig. 5a or elsewhere on the ground (Ren et al. 2004b, 2008, 2009) as shown in Fig. 5b. Conaire et al. (2009) used the infrastructure which included UbiSense spatial location system and nine cameras, with pan, tilt and zoom capability, positioned around the court as shown in Fig. 5c. Ohno et al. (2000) placed four cameras on each side of the soccer stadium as shown in Fig. 5d.

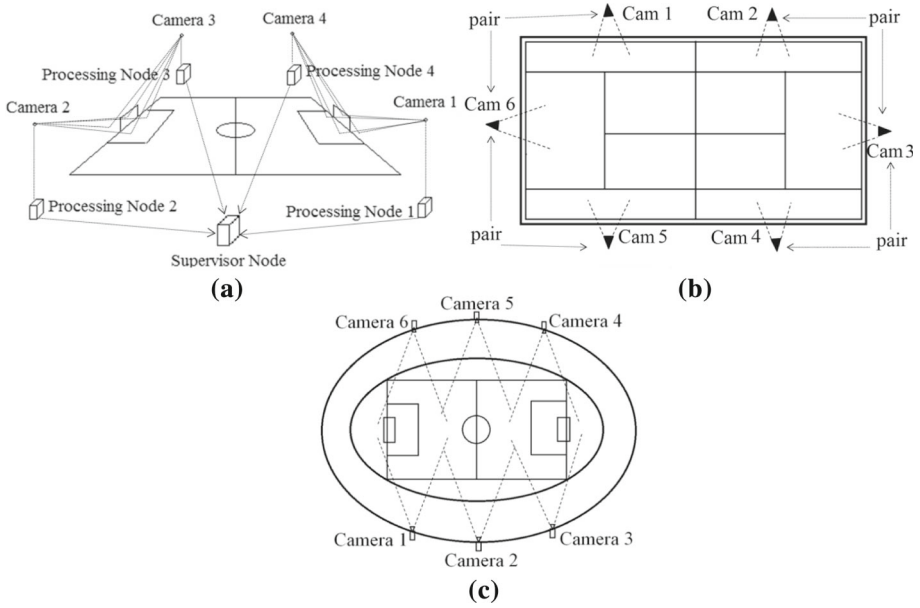


Fig. 4 Four to Seven camera system. **a** Four cameras placed on each side of goal posts (D’Orazio et al. 2009a,b). **b** Four pairs are formed by six cameras. **c** Six cameras placed on opposite sides of stadium (Leo et al. 2008)

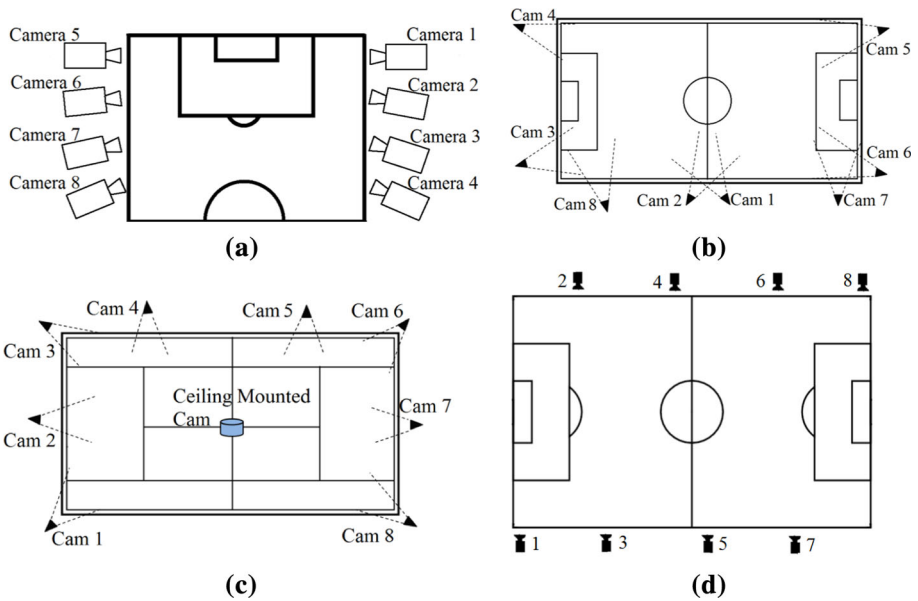


Fig. 5 More than Seven camera system. **a** Eight cameras aiming at goalpost (Iwase and Saito 2003). **b** Eight cameras covering the whole ground (Ren et al. 2009). **c** Eight PTZ cameras placed around the court and one ceiling mounted (Conaire et al. 2009). **d** Four cameras on each side (Ohno et al. 2000)

Single cameras are used where the field of view is small and fixed viz., goalpost or if the court is small enough to be captured from a single position as in Tennis. Single high-speed camera tracks all three motions; the rotational, translational and spins of a ball. External robotic mechanism gives 3D movement to single camera to keep the ball in its field of view. However, single camera system could not solve the problem of occlusion of ball by players due to its limited viewing angle. Multiple camera system could deal with occlusion problem by giving clearer views from multiple angles. Such systems cover every nuke and corner of the stadium. Modern day sports entertainment needs 3D localization of the ball and players as well as 180° shot from multiple cameras.

2.2 Field modeling and registration

The standard dimensions of the court model, lines and circles on the court, net in tennis, ring in basketball sport, etc. are available in the rulebook of the respective game and used as domain knowledge. The dimensions prove to be of much help in marking the field of interest and modeling the sport court, ground or field. The model and the crossing points of lines form the feature points that are required for calibrating the camera(s). In order to increase the number of intersection points, virtual intersection points (Tong et al. 2011) are also employed along with real intersection points.

The registration stage involves extraction of feature points from the ground lines and their crossings. General methods incorporating the use of Hough-transform (Chen et al. 2012; Farin et al. 2004; Miyamori and Iisaku 2000; Tong et al. 2011; Yu et al. 2009) or Canny edge detector (Ren et al. 2009) for feature point extraction by line detection and/or custom made ellipse detection (Yu et al. 2003d, 2004a) for tangential feature point extraction are used to find feature points. Feature points are also searched in other planes like tennis net (Chen et al. 2012; Miyamori and Iisaku 2000; Yu et al. 2009) using similar procedure. It is followed by tuning the homography matrix (or camera parameter refinement (Farin et al. 2004)) and updating for more accurate feature points. After obtaining all the straight lines in different planes, some reference like net is used for fitting the straight lines of the image with the model (Yu et al. 2009). The lines may cause false alarms which are removed using masks made from Field of View (FOV) of each camera in a multi-camera system (Ren et al. 2009). The Hough transform for line detection suffers from a disadvantage of being slow. To add to the complexity, the lines in Broadcast Soccer Video (BSV) are long and sparse. A solution was provided in Yu et al. (2004c) in the form of a grid based algorithm for line detection in goal-mouth scene. The system designed in Farin et al. (2004) required 1 s for initialization which was not sufficient for real-time applications which was modified by Farin and Han (2005) and increased the speed to 30 ms for first frame and 6 ms during court tracking using RANSAC (Fischler and Bolles 1981) based line detector for line parameter estimation instead of Hough transform. On the lines of Farin and Han (2005), Chang et al. (2009) also modified the approach of Farin et al. (2004) and made it applicable for broadcast basketball videos (BBV).

2.3 Camera model and calibration

Camera calibration consists of model estimation for an un-calibrated camera. The main objective of the method consists of finding the external (extrinsic) and internal (intrinsic) parameters of the camera. External parameters of camera include its position and orientation with respect to the world coordinate system whereas the internal parameters include the image center, distortion coefficients and focal length. The calibrated camera video signal

provides an accurate and consistent mapping between the real world objects' position and dimensions and their graphical representation in the video frames.

2.3.1 Ideal pinhole camera model

The video acquisition process always employs image acquisition to grab video frames for research work. Thus, we describe the basic static camera model for image acquisition for the sake of completeness. The model defines the geometric relation between a 3D point in real world coordinate system and its corresponding 2D projection on the image plane. A real world point $(X, Y, Z)^T$ in 3D homogeneous coordinate space is $P = [X, Y, Z, 1]^T$ after adding a fourth auxiliary coordinate 1. Its corresponding 2D point $(x, y)^T$ on the image is $p = [x, y, 1]^T$ with the third auxiliary coordinate 1. The points in 3D and 2D coordinates are related as in Eq. (1).

$$\lambda p = \mathbf{H}P = \mathbf{K}[\mathbf{R}|\mathbf{T}]P \tag{1}$$

With $\mathbf{K} = \begin{bmatrix} f_x & c & o_x \\ 0 & f_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$ $\mathbf{R} = [r_0 \ r_1 \ r_2] = \begin{bmatrix} r_{00} & r_{01} & r_{02} \\ r_{10} & r_{11} & r_{12} \\ r_{20} & r_{21} & r_{22} \end{bmatrix}$, $\mathbf{T} = [t] =$

$[t_x \ t_y \ t_z]^T$ where, \mathbf{H} is the perspective projection matrix, \mathbf{K} is the camera intrinsic parameter matrix which include (f_x, f_y) as focal lengths on x and y axis in image plane, (o_x, o_y) as principal points which denote the intersection of the optical axis of the camera lens and the image plane and c is the distortion coefficient of the two axes of the image. \mathbf{R} and \mathbf{T} denote the rotation matrix and transition vector as the extrinsic parameters. λ is the homogenous scaling factor. Equation (1) modifies into $p \approx \mathbf{H}P$ if scale factor is unity. There are total eleven camera parameters that include five intrinsic and six extrinsic parameters used for achieving camera calibration.

At least six pairs of non-coplanar points are required between the world coordinates and the image coordinates for estimating all eleven parameters (Chen et al. 2009). By assuming the height in the world coordinate system as constant and scale factor as unity, Eq. (1) can be modified into Eq. (2) where r_2 in rotation matrix and Z component are made zero (Hartley and Zisserman 2003).

$$p = \mathbf{H}P = \mathbf{K}[\mathbf{R}|\mathbf{T}]P = \mathbf{K} [r_0 \ r_1 \ r_2 \ t] \begin{bmatrix} X \\ Y \\ Z = 0 \\ 1 \end{bmatrix} \approx \mathbf{K} [r_0 \ r_1 \ t] \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{2}$$

The perspective projection matrix size reduces from 3×4 to 3×3 . The number of camera parameters also reduces from eleven to eight. The computation of eight camera parameters is thus possible using only four pairs of coplanar points (Chang et al. 2009; Liu et al. 2006; Mauthner et al. 2007; Ohno et al. 1999; Yu et al. 2004c). If the image's \mathbf{H} is known and is invertible then its inverse can be used to calculate a point in 3D coordinate frame (Liang et al. 2007; Liu et al. 2006) if its corresponding point in 2D coordinate system viz., image, is available.

The method covered so far in this sub-section is direct homography estimation. It is used only if sufficient key-points are available. If the number of key-points in the current frame is not sufficient then indirect method of homography estimation is used (Tong et al. 2011). A few camera calibration methods (Chen et al. 2012; Yu et al. 2004c) used direct linear transform (DLT) (Hartley and Zisserman 2003).

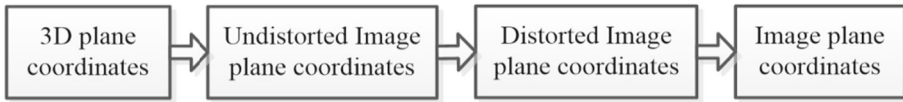


Fig. 6 Basic block diagram of Tsai method of camera calibration

2.3.2 Tsai camera calibration model

This is one of the most popular camera calibration methods since it is suitable for application to wide area as it can deal with coplanar as well as non-coplanar points. It also facilitates separate calibration of internal and external parameters. This process is mostly suitable for cases with known internal parameters and unknown external parameters. This method involves transformation from 3D world coordinates to undistorted image plane coordinates, then to distorted image plane coordinates and finally to image coordinates (Kumar et al. 2011; Li et al. 2005; Pingali et al. 2000, 2002) as shown in Fig. 6.

Along with the intrinsic and extrinsic parameters, distortion factors could also be calculated as in Choi and Seo (2011). The object’s measurement in the world coordinate could be calculated using the pixel dimensions in image as two fixed intrinsic constants (Ren et al. 2008). The manual mapping of the corner points in 2D image model and those on 3D pitch model are inputs to the Tsai’s calibration. Tsai’s camera calibration algorithm is of less use for the cases of ceiling mounted cameras, as it requires 3D coordinates for calibration unlike the 2D plane of the observed field.

2.3.3 Using geometric relations and multi-camera homography

2D projective transformation is a linear transformation using a homogenous 3-vector transformation H represented by a non-singular 3×4 matrix as in Eq. (3).

$$\begin{aligned}
 x = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &= \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\
 &= \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = HX \quad (3)
 \end{aligned}$$

where, H is a homogenous matrix. Let the inhomogeneous coordinates of a pair of matching points X and x in the world and image plane is given by (X, Y, Z) and (u, v) respectively. The projective transformation in inhomogeneous form is written as in Eq. (4).

$$u = \frac{x_1}{x_3} = \frac{h_{11}X + h_{12}Y + h_{13}Z + h_{14}}{h_{31}X + h_{32}Y + h_{33}Z + h_{34}}, \quad v = \frac{x_2}{x_3} = \frac{h_{21}X + h_{22}Y + h_{23}Z + h_{24}}{h_{31}X + h_{32}Y + h_{33}Z + h_{34}} \quad (4)$$

Thus we get the linear equations, Eq. (4), in terms of H . Four point correspondences leading to eight such equations are sufficient to solve H . This method is usually applied in the system where two (Choi and Seo 2004; D’Orazio et al. 2009a,b) or more cameras [six in case of D’Orazio et al. (2009a,b) and eight in case of Iwase and Saito (2003)] are used and homography is to be computed between them. The feature points may be derived automatically (Zhu et al. 2008) or supplied manually (Choi and Seo 2004). Such reference points are usually dependent on the scene the cameras are capturing. In case of football, if

the scene happens to be the goalmouth region then the penalty area or goal area lines (Iwase and Saito 2003) may be used and if it happens to be the center of the field then the tangential points on the ellipse are used (Yu et al. 2003d, 2004a). In case of two cameras with focal points close to each other, their stitched image is used for computation purpose (Choi and Seo 2004). Finally, this method computes the homography between the stitched image and one of the two images.

2.3.4 Toolboxes

Camera parameters may also be calibrated automatically using OpenCV camera calibration toolbox as in Conaire et al. (2009) and Matlab camera calibration toolbox as in Kumar et al. (2011). Intrinsic and extrinsic parameters may be calibrated separately as in Kelly et al. (2010) which used Matlab toolbox to calculate intrinsic parameters and OpenCV toolbox to calculate extrinsic ones. Kumar et al. (2011) used Matlab calibration toolbox for calibrating every single camera. They obtained 3D points using the 2D point triangulation and back-projection methods. The authors used weighted averaging method to avoid camera calibration errors and 2D tracker errors.

In case of a single static camera system with fixed and limited area, ideal pinhole camera system may prove beneficial. It is the simplest case of homography estimation. However, if the system has more than one camera; as is required in most of the games and sports the multi-camera homography method is suitable as compared to the ideal single pinhole camera system. Tsai's camera calibration method is preferred when wide area is to be covered. It can also deal with coplanar and non-coplanar key-points. In Tsai's method, it is possible to calibrate internal and external parameters separately. However, Tsai's method is less suitable for ceiling mounted camera system, as it requires 3D coordinates for calibration while the field is a 2D plane for a ceiling mounted camera.

2.4 Ground truth

In tracking, ground truth refers to either actual 2D or 3D position or contour of the object in every frame of the video sequence (Zhang et al. 2010). The overall ground truth data of the video sequence is a standard for quantitative evaluation of any tracking algorithm. Capturing the ground truth data becomes an important step after camera calibration and registration. Usually, an operator obtains the data manually by observing every frame and selecting the region or points. This happens to be a cumbersome and time-consuming task. Software like Video Performance Evaluation Resource (ViPER) (Chakraborty and Meher 2013b; Li et al. 2005), Context Aware Vision using image based Active Recognition (CAVIAR) (Li et al. 2005) and Open Development for Video Surveillance (ODViS) (Li et al. 2005) provide an interactive framework for defining ground truth. Data generated by a tracking system is compared to the ground truth for performance evaluation or rule infringements detection.

2.5 Foreground extraction

The term 'Foreground' refers to a set of pixel intensities that drastically change their positions in current frame with respect to their positions in previous frame. The set of such pixel intensities form an object that appears moving in every next frame. On the contrary, 'Background' refers to a set of pixel intensity values that either do not change their positions or change their positions by a few pixels per frame. The process of foreground extraction assumes that the trackable object is a part of foreground and without separating it from the background; detec-

tion, and eventually tracking, will not be feasible. Thus, foreground detection or background detection both aim at object (target) detection.

Since the target object is ball, it is evident that players will play with it on a ground or court and spectators will watch the game. The spectators are the biggest source of introduced noise and distortion, if they are visible in a video sequence. Thus, some approaches (Barnard and Odobez 2004) have introduced playfield extraction as the first step in foreground extraction by building a mask for pruning the noisy stands region. Further, the approaches extract the non-field regions inside the playfield as foreground. The process generally uses dominant color of the playfield sometimes coupled with histogram back-projection (Dearden et al. 2006) to build a mask. However, masks built from fusion of geometry based masks and dominant color based masks are also used (Ren et al. 2008). The court masks can also be generated using RANSAC based method or Hough Transform based method (Naidoo and Tapamo 2006) of court boundary line detection accompanied with dominant color based method (Chang et al. 2009). Sometimes color alone is not sufficient so multiple cues including color, motion (correlation between successive frames) and shapes are used for forming the playfield segmentor (Xing et al. 2011). However, during long shot frames, smaller ball size may hamper the ball detection process. Thus, a few methods suggested pre-processing of frames in advance using Saliency Map (SM) and later constructing the Saliency Binary Map (SBM) for facilitating the detection process (Chakraborty and Meher 2013a). In case of BSV, the above approaches filter out the frames containing the field view. The color ranges of the field, lines, ball and player uniforms are provided as input to the algorithm which does the sorting task (Yu et al. 2006).

Playfield extraction may not always be included in the foreground extraction process. Sometimes, background subtraction replaces foreground extraction. It requires building an adaptive background model from which foreground objects are extracted. The model can be built either beforehand or during runtime. For an object to be considered foreground two cues are generally used viz., motion and color.

Using motion as feature includes generation of the background model based on per-pixel Gaussian mixture model (GMM) (D’Orazio et al. 2009a,b; Li et al. 2005; Ren et al. 2008, 2009). The model must be updated in order to overcome errors in initial background registration due to lighting variations. Some algorithms like running average may be used for updating the model efficiently (Ren et al. 2008). Background subtraction algorithms like recursive approximate median filtering (Chakraborty and Meher 2012b; Ekinici and Gokmen 2008), successive frame differencing (Chakraborty and Meher 2011, 2013b; Chen and Wang 2007; Chen et al. 2007; Conaire et al. 2009; Ishii et al. 2007; Kokaram et al. 2005; Zhang et al. 2010) etc. use sequence of successive frames to distinguish moving pixels from the static ones.

Using dominant color as feature, statistical clustering approach like GMM was used to build the playfield color model (Barnard and Odobez 2004; Choi and Seo 2004; Kasiri-Bidhendi and Safabakhsh 2009; Liang et al. 2007; Zhu et al. 2008) based on field color. Dominant color pixels may be used to build a court binary mask (Chang et al. 2009; Hossein-Khani et al. 2011; Hu et al. 2011; Kim and Kim 2009; Tong et al. 2004) whereas the non-dominant pixels formed the player and ball candidates. The model needs to be updated periodically in order to counter the lighting variations (Kelly et al. 2010). The process of model adaptation or parameter estimation may be carried out by maximum a posteriori (MAP) or expectation-maximization (EM) (Liang et al. 2007). The process of foreground extraction can be implemented by forming lookup tables of the ball color (Liu et al. 2006) and comparing each pixel with it using some distance measure.

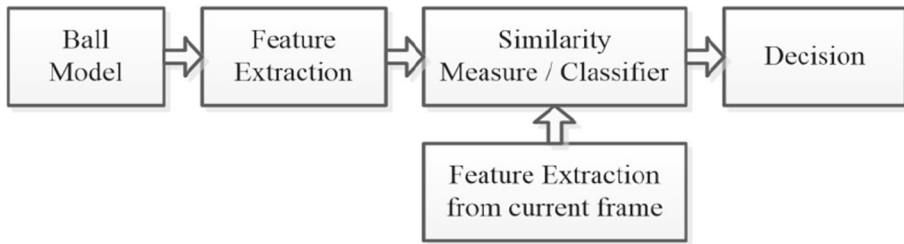


Fig. 7 A general ball detection system

The resulting image is converted into binary form so that the ball regions are vividly seen. Logical operations (Ishii et al. 2007; Kokaram et al. 2005; Pingali et al. 2000, 2002; Ren et al. 2008) may also be helpful in such cases. To get rid of noise and clutter, morphological operations like opening (Batz et al. 2009; Chakraborty and Meher 2012a, 2013a, b; Chen et al. 2012; Kittler et al. 2005; Zaveri et al. 2004), closing (Chakraborty and Meher 2012b; Ekinci and Gokmen 2008; Hossein-Khani et al. 2011; Liu et al. 2006; Pingali et al. 2002; Ren et al. 2008; Shum and Komura 2005), top-hat (Ekinci and Gokmen 2008), etc. may be applied on the binary images. Thus, the number of false positives is reduced using morphological operations which are sometimes assisted by region growing processes (Liang et al. 2007; Yu et al. 2004b). The obtained blob regions are labeled in order to provide them identity. Thus, connected component labeling (Figueroa et al. 2004; Kim and Kim 2009) is performed and minimum bounding rectangle (MBR) may be fit (Naidoo and Tapamo 2006).

Like players, tracking of a ball is important due to the fact that intention of the players cannot always be judged from their movements only (Ohno et al. 1999).

3 Features and methods for ball detection

This stage involves the discussion of various features and methods used to detect the ball from the blobs obtained after the foreground extraction stage. A basic system for ball detection is shown in Fig. 7.

The initial stage involves modeling of the ball region, which is followed by feature extraction from samples and storing (or mapping) it for future use. This modeling of ball is carried out from different distances, under different illuminations and from different angles in case of non-spherical balls. This even may be required in case of multi-colored or multi-pattern balls. It is also known as training stage in some works. Either a similarity measure or a classifier is chosen which helps to identify the regions into Ball or non-Ball by obtaining the features from the blobs of the sequence. This process is sometimes referred to as testing phase as well. The next section presents the features and methods used by researchers for detecting ball from the blob regions.

3.1 Color

In most of the sports played worldwide, the balls have a distinct color compared to playfield or ground. In order to detect a fast moving ball its color proves to be one of the important cues. The detected blobs are searched for the specific color pixel clusters and if found, may be classified as a ball based on other criteria like shape. The ball pixels are obtained manually beforehand during training. The ball color sieve may be designed to filter objects with too

Table 2 Summary of performance measures of references where color was used as a feature for ball detection

References	Sports (%)	Avg. accuracy (%)	Avg. precision (%)	Avg. recall (%)
Naidoo and Tapamo (2006)	Soccer	70.5	–	–
Yu et al. (2007b)	Soccer	83.04	–	–
Hossein-Khani et al. (2011)	Soccer	–	80.05	83.46
Liu et al. (2006)	Soccer	66.72	92.6	61.8

few or no ball color pixels (Ekinici and Gokmen 2008; Yu et al. 2003b, 2004b, 2005, 2007b, 2009). A general condition for segmenting ball candidate pixels in normalized color space is presented in Eq. (5) (Huang et al. 2008; Liang et al. 2005; Liu et al. 2006; Liang et al. 2007).

$$B(x, y) = \begin{cases} 1, & \text{if } (r(x, y) - 1/3)^2 + (b(x, y) - 1/3)^2 \leq a^2 \wedge I(x, y) \geq b \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

Here, B —binary image, (x, y) —pixel locations, $r(x, y)$ and $b(x, y)$ denote normalized red and blue components respectively, $I(x, y)$ —luminance value. Empirical values of a and b are selected heuristically using the color of the ball during training. The detection process for a blob also depends on the ratio of the ball color pixel to the total number of blob pixels (Ren et al. 2008; Yu et al. 2006). A template histogram (Wang et al. 2014a) formed by concatenation of histograms of HSV channels separately obtained during training have been used for comparison with the current histograms. Table 2 summarizes ball detection results of a few research works.

However, color feature of a ball is highly affected by illumination, sunlight, floodlight, weather conditions, type of lighting in indoor court, ball speed, player uniform, background with similar color objects, too bright or too dark background, etc. which makes it unreliable for detection purpose.

3.2 Size

It is a flagship feature that needs to be estimated in pixels, from actual ball’s dimensions in real-world units, before searching in frames. Size of an object is determined by the total number of pixels it is having which is decided either according to its 8—connectedness (Ekinici and Gokmen 2008) or the size of the bounding box. Its range sets a limitation for the length and width of the object’s MBR. Average player’s height could be used to determine ball size range for each frame F using a function $R, R(F) = [R_{\min}, R_{\max}] = [h/7, h/3]$, where, h is the average height of all players in F (Chen et al. 2009; Yu et al. 2003a). Ball size estimation from player’s height is less accurate than using center ellipse, goalmouth (Yu et al. 2003b, 2004c, 2006), court lines (Chen et al. 2012) or tennis net (Yu et al. 2004b). The ball size range may be given as in Eq. (6).

$$s(i, j) = [b(i, j) \cdot (1 - \Delta_1), b(i, j) \cdot (1 - \Delta_2)] \quad (6)$$

The result replaces the ball size $b(i, j)$, where, Δ_1 and Δ_2 are the extensions for error tolerance. As the ball size may vary from frame to frame (Chakraborty and Meher 2011, 2012a, b, 2013a) adopted suitable pixel ranges for detecting a ball candidate in a video of varying frame sizes. Homography could also be used to estimate the ball size (Yu et al. 2007b, 2009). Ishii et al. (2007) implemented a real-time ball detection which took an average time

of 3.08 ms per frame by the ball centered search. [Chen et al. \(2007\)](#) obtained an accuracy of 92.43%.

3.3 Shape

Shape is another important feature for ball detection. Its measurement and representation is carried out either by direct detection or removal of non-ball objects. It is categorized into many sub-features. A brief compilation of different shape sub-features, their mathematical representations, their typical values or ranges have been presented in [Table 3](#). [Table 3](#) also presents different performance measures achieved and published so far using these shape features by different algorithms followed by their shortcomings if any.

Shape of the ball is subject to change depending on type of camera, its framerate, ball speed, background, etc. In spite of these problems, shape is the most used feature for ball detection.

3.4 Position/proximity to expected position

Position/proximity is used in case of multiple detections. The ball candidate is supposed to be found in the expected position only ([Pingali et al. 2000](#); [Teachabarikiti et al. 2010](#)) and the ball candidates that are far from the hitting position must be discarded ([Yu et al. 2004b](#)). Accurate automatic detection of hitting direction/position is a major issue of concern. However, the position feature plays an important role in detecting true ball candidates. Since the camera always follow the ball it is most likely to appear in 2/3rd width in the center and 2/3rd height at the bottom of a frame ([Zhu et al. 2008](#)).

3.5 Template matching

This method of ball detection is required to have a ball-like template formed beforehand. The process of template matching (TM) is performed at the player contour and is repeated for a straight-line path of the ball moving radially away from the player and at a predicted position. TM is generally performed using the normalized cross correlation between the template obtained from the first frame and the detected ball candidates ([Ariki et al. 2008](#)). The gray correlation coefficient ([Tong et al. 2004](#)) and the correlation measure ([Leo et al. 2008](#)) are on the same lines as in [Ariki et al. \(2008\)](#). An adaptive TM method ([Ishii et al. 2007](#)) is also used which is based on ball movement and reduces the search area as much as possible. Probability map based TM ([D’Orazio et al. 2009a,b](#)) uses past information of ball motion. The positive and negative ball image templates may be used to train a support vector machine (SVM) classifier for ball detection ([Ancona et al. 2003](#)). However, TM may give false positives if objects with similar appearance like player’s head, socks, shoes, ball-like pattern on jersey, scoreboard, stands, etc. are present in the current frame. Frequent update of template is a solution to reduce the number of false positives.

3.6 Circle and ellipse detection

This subsection discusses a few algorithms dedicated to circle detection viz., circle Hough transform (CHT), modified circle Hough transform (MCHT) and ellipse fitting.

CHT extends the basic idea of Hough transform from line detection to circle detection. The original CHT is robust to occlusions and noise but suffers from disadvantage of computational complexity. As per the comparison done in [Janssen et al. \(2012\)](#), the speed of modified CHT

Table 3 Summary of shape features, their mathematical representation, published performance and their limitations

Shape features	Representation	Valid range/value	Avg. accuracy	Shortcomings
Aspect ratio	R_{h-w}, R_{w-h} (Yu et al. 2003b, 2004b,c) where, R_{h-w} and R_{w-h} is height to width and width to height ratio of object resp. $[R_{h-w}, R_{w-h}]$ (Chakraborty and Meher 2011, 2012a, 2013a; Chen et al. 2007, 2009; Mitra et al. 2009; Yu et al. 2009)	Less than a statistically decided threshold Within the range	- 96.22% (Chakraborty and Meher 2011), 94.77% (Chakraborty and Meher 2012a), 94.65% (Chakraborty and Meher 2013a) ^a	Can detect non-circular objects as well
Form factor/roundness	Ratio of length and width of object's MBR (Ekinci and Gokmen 2008; Liang et al. 2005; Liu et al. 2006; Liang et al. 2007) Form factor, $F = \ P\ ^2 / (4\pi A)$ where, P is the perimeter and A is the area of the object (Huang et al. 2008; Zhang et al. 2010) Roundness= $(4\pi A) / \ P\ ^2$ (Kasiri-Bidhendi and Safabakhsh 2009; Wong and Dooley 2010)	- Close to 1 (Tong et al. 2004), heuristic (Naidoo and Tapamo 2006) For circular object, $0 < \text{Roundness} < 1$ For circle, Roundness = 1 For circular object, C is very large, For circle, $C = \infty$	- 91.8% (Chakraborty and Meher 2012b) ^a	- Speed of 11 spf, misdetection, tracking failure (Kasiri-Bidhendi and Safabakhsh 2009)
	Circularity (Chen and Zhang 2006; Tong et al. 2004), $C = \frac{\mu_R}{\sigma_R}$ where, μ_R and σ_R are mean and variance of the distance of all K pixels on contour from centroid (\bar{x}, \bar{y}) of the object. Mean and variance are given by: $\mu_R = \frac{1}{K} \sum_{i=0}^{K-1} \ [(x_i, y_i) - (\bar{x}, \bar{y})]\ $ and $\sigma_R = \frac{1}{K} \sum_{i=0}^{K-1} [\ (x_i, y_i) - (\bar{x}, \bar{y})\ - \mu_R]^2$			

Table 3 continued

Shape features	Representation	Valid range/value	Avg. accuracy	Shortcomings
Eccentricity	Eccentricity, $E = D_L / D_S$ where, D_L and D_S are the longest and shortest diameters of the MBR of the region under consideration respectively (Chakraborty and Meher 2013b; Naidoo and Tapamo 2006; Tong et al. 2004).	Close to 1	–	Can detect non-circular objects as well
Degree of fullness (Chen et al. 2007)/compaction filter	Degree of compactness, $D_c = \frac{\text{Object_size}}{\text{Bounding_Box_Area}}$	More than 50% (Chakraborty and Meher 2011, 2012a)	71.28% (Chen et al. 2007), 71.85% (Chen et al. 2012)	–
	Ratio of area of the object region to that of MBR (Liang et al. 2005, 2007; Liu et al. 2006; Pei et al. 2009) or Minimum bounding polygon (MBP)	–	–	–

Table 3 continued

Shape features	Representation	Valid range/value	Avg. accuracy	Shortcomings
Rounded upper contour (RUC) (Wong and Dooley 2010)	$RUC = \begin{cases} 1 & \text{if } ERUC < IRUC \\ 0 & \text{if } ERUC \geq IRUC \end{cases}$ <p>where, $ERUC$ is an objective (error) function given below and $IRUC$ is a preset threshold.</p> $ERUC = \frac{\sum_{i=1}^n d_i - r }{N}, d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}$ <p>where, (x_i, y_i) is an upper contour pixel, i is pixel index, N is the total number of pixels, r is the radius and (x_c, y_c) is the center obtained by solving the circle equation for (x_i, y_i) as in Fig. 8. d_i is the distance between center and pixel i. The upper and lower contour of a ball is visible depending upon whether it is held by the player from the upper side or lower side</p>	-	-	-

^a The ball detection algorithm obtained an average precision of 89.62%.

^b The ball detection algorithm obtained an average precision of 97.12%

techniques depend on number of edge pixels and the parameter space size. In case of Football as an object, its number of edge pixels is very small and its size does not change much. Due to these reasons, the original CHT was found suitable for real-time purposes compared to modified CHT. In Wang et al. (2014a), color histogram based TM was used to find ball candidates. CHT was applied on the obtained ball candidate regions by Pallavi et al. (2008).

MCHT involves the use of directional CHT with simple convolutions (D'Orazio et al. 2004, 2009a,b) or using improved Circle response (Birbach and Frese 2009; Birbach et al. 2011; Cigliano et al. 2015). The prior method requires knowledge of ball appearance (illumination conditions and ball dimensions) to prepare masks for the ball to be detected. The method is advantageous in terms of high detection rates during full visibility of ball and low computational cost. The main drawback of the method was its failure in detecting occlusions and ball's absence in the frame. Finally, a trained neural network (NN) was used to classify the candidate into ball and no ball. The later method is illumination invariant and does not need tuning parameters or hard thresholds. D'Orazio et al. (2004) implemented a modified Atherton algorithm which gave 97% of ball detection under natural light conditions with a processing speed of 0.3 s/frame. Researchers Birbach and Frese (2009) yield 69.4% of ball detection for a 417 stereo frame sequence with visible balls forming 32 trajectories. However, CHT requires prior knowledge of the ball radius, which is to be detected. Thus, it cannot be used in multiple shot types without modifying the radius. CHT with NN fails to detect the ball when non-ball objects look like the ball.

It is observed that if the ball moves very fast or if the frame rate is less; it appears as a blurred ellipse instead of circular. Generally, the ellipse is fitted on the blob, with its major and minor axes crossed through the ball center, according to some means like least square criterion (Yan et al. 2005). Quality measure of the fit can also be extracted as in Christmas et al. (2007). Finally, a trained classifier like Adaboost (Kittler et al. 2005) or SVM (Christmas et al. 2007) is used to classify the blobs into ball or non-ball. A few approaches like Yan et al. (2005) used the mean absolute difference between the gradient and the normal direction at all sample points.

3.7 Motion

Once the player region is detected, the ball motion region is searched in radial direction from the players at the estimated position (Ohno et al. 2000, 1999). Its representation was provided in Wong and Dooley (2010) and given in Eq. (7).

$$M = \begin{cases} 1, & \text{if } O_{diff} \geq t_M \\ 0, & \text{if } O_{diff} < t_M \end{cases} \quad (7)$$

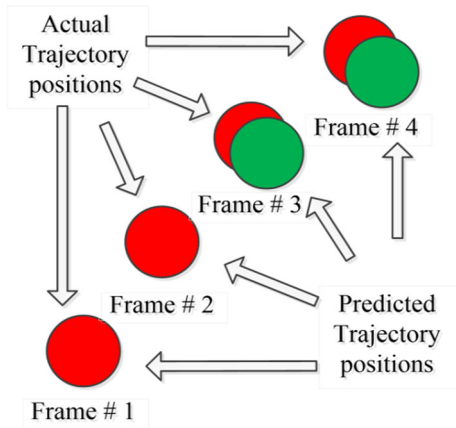
where, M represents the condition as to whether the ball candidate is in motion or not, t_M is the preset threshold and O_{diff} is the Euclidean distance between the centroid of the blob in the current and the previous frame. Pallavi et al. (2008) used motion based detection for non-ball filtering in medium shots in BSV.

3.8 Trajectory

The trajectory-based detection involves the difference between the predicted and actual trajectory. It is a condition represented by T , as to whether the ball candidate follows the predicted trajectory or not. It is proposed in Wong and Dooley (2010) and given in Eq. (8).

$$T = \begin{cases} 1, & \text{if } L_{diff} < t_T \\ 0, & \text{if } L_{diff} \geq t_T \end{cases} \quad (8)$$

Fig. 8 Red balls show the predicted trajectory and green balls show the actual one



where, t_T is the preset threshold and L_{diff} is the Euclidean distance between the actual and predicted locations of the blob. The method is shown in Fig. 8 where, circles represent ball in a trajectory at the successive frames.

In frames 3 and 4, the red (upper) circle represents actual ball trajectory whereas the green (lower) circles are part of predicted ball trajectory. The predicted locations are the linear extrapolation of the blob position in the previous frames 1 and 2. Ideally, the actual trajectory position and the predicted trajectory positions should exactly match. Pallavi et al. (2008) achieved non-ball filtering in long ball shots using trajectory obtained by dynamic programming. They achieved ball detection at an average recall rate of 94.2% and average precision of 90.72%. This algorithm does not need prior pre-processing of the video and is efficient even with poor playfield conditions. However, they could not detect the ball when it is partially visible in medium shots. In addition, they could only estimate the routes for medium and long shots and not for the whole match.

3.9 Gradient directions

After ellipse fitting is applied on the blobs, if the gradient is assumed to be normal to the ellipse and positive to the direction of the center of the ellipse then the blob may be a ball (Kittler et al. 2005). Based upon this, the blobs are classified using Adaboost classifier. Blobs are classified into ball or no-ball using the gradient directions taken at the blob boundaries (Almajai et al. 2012; Huang et al. 2011).

3.10 Training based detector

Instead of using single feature based ball detection, a multi-feature Bayesian classifier may be used for classifying the objects into ball and non-ball. The generally used features are of two types, viz. appearance features and distance features. Some appearance features (Yu et al. 2003b, 2004b, c) may include surrounding of the candidate, color, shape, size, circularity, etc. whereas distance features (Yu et al. 2006, 2007b, 2009) include separation. An SVM classifier, trained for blob color and shape features has been used for classification in Christmas et al. (2007).

3.11 Speed/velocity

Speed is an important cue for ball detection and removal of false candidates (Ishii et al. 2007). This is because ball is the fastest object on the ground and its image will have motion blur proportional to its speed. Suppose that a ball, whose diameter is 22 cm, moves with a maximum speed of 100 km/h in a frame whose rate is 30 fps, then the ball will move around 90 cm on ground during a single frame interval. Obviously, the ball moves a distance several times the apparent ball diameter in a frame.

4 Ball tracking techniques

Main purpose of a ball-tracking algorithm is to establish a correspondence between detected ball candidates across adjacent frames to form a continuous trajectory. Ball tracking plays another important role of alleviating discrepancies like misdetection and hidden or missing ball localization. Following subsections elaborate on various ball tracking approaches and their implementation followed by their strength and weaknesses.

4.1 Kalman filter

This approach deals with estimating optimal state of a ball in the current frame if the initial (previous) state and noise are available. Kalman filter (KF) works well for linear systems with Gaussian distributed states and noise. However, for non-Gaussian object states and for non-linear systems, particle filter (PF) is preferred for state estimation.

Most of the researchers, in medieval era, employed KF to estimate ball velocity for predicting ball position i.e. detection in each frame. However, ball size, color, longevity have also been modeled for improving the ball detection process (Ren et al. 2006). They employed occlusion reasoning and tracking back for obtaining the ball occluded by players as well as for false alarm removal. Robustness and continuity of the ball tracking system was improved using hysteresis based thresholding by Ren et al. (2006). KF predicted vector was also used for improving the robustness and continuity of ball tracking in Ishii et al. (2007). Position, velocity and acceleration of the ball are modeled in 2D or 3D space in Ishii et al. (2007). Chen and Zhang (2006) incorporated incremental Bayesian algorithm (EM) in association with KF to update dynamic motion and appearance parameters of the ball. Ren et al. (2009) used multi-camera based approach similar to Ren et al. (2006) for ball tracking in 3D space.

During ball tracking, if players occlude the ball for a long time, conventional KF makes a poor ball prediction that increases error. Dynamic KF (DKF) has been used for reducing the error in such dynamic conditions using occluding player information and controlling velocity of the state vector (Kim and Kim 2009). However, the proposed framework shows poor results when the players are crowded on the ground around the ball. During any sport, the ball appearance keeps on changing due to various environmental factors. Large change in ball appearance causes KF to fail due to the changed reference statistical appearance model. The changes may be incorporated in observations of KF at regular interval by adaptively updating the ball template (Liang et al. 2005, 2007; Liu et al. 2006). KF predicts location of the ball in the next frame and filters the tracking result in the current frame. The algorithm runs fast owing to the fact that only a small portion of each frame is processed in most of the cases when the ball is detected around its expected position.

We can broadly classify ball motion in sports into linear and non-linear motion. Single linear filter is not able to track the ball in that condition. A few researchers used multiple-

filter banks that took care of both the motions by tracking the ball in maneuvering as well as non-maneuvering state with fast switching method (Zaveri et al. 2004). They did not use any apriori information regarding the object dynamics. The system was robust for noisy background and changing ball size due to camera zooming effect. However, the datasets used were too small i.e. of mere 100 and 32 frames. Desai et al. (2005) implemented a similar multi-filter bank tracking algorithm on Xilinx Virtex-II Pro FPGA, but no quantitative results were presented. A single motion track using KF cannot model the discontinuity of ball motion due to a bounce or a hit. This problem can be solved by generating Multiple hypotheses i.e. one track for each motion model (each new field) to cope with the discontinuities (Kittler et al. 2005).

KF is a basic tracker suitable for tracking non-occluding objects moving linearly, slowly and with constant acceleration. Thus, tracking a high velocity ball with changing appearance is not possible with KF. Though a few methods claim to achieve good results with KF, they use either smaller datasets with low-resolution frames or deliberately avoided videos with conditions like cluttered background, complex trajectory, occlusion, etc. Table 4 presents a summary of all KF based ball-tracking methods published so far in different sports, validating databases, obtained results, strengths and shortcomings.

4.2 Particle filter

For non-linear systems, many extensions of KF have been developed and applied for ball tracking. However, when a strong non-linearity i.e. high ball speed, acceleration and spin occurs, KF and its extensions fail to give precise approximations of mean and covariance matrices and this degrades the tracker performance drastically. In such a scenario, Sequential Monte Carlo or Particle filter (PF) (Arulampalam et al. 2002) is suitable for approximating the posterior probability density function of the state. The algorithm propagates a weighted set of particles that approximate a density function. PFs are neither bound for utilization in presence of Gaussian noise nor bound by linearity of the systems. Thus, they provide flexible tracking frameworks. PF is suitable for solving the problem of ball tracking as balls have varying sizes, non-constant velocity, prone to occlusion due to players, objects and background with similar appearances. Generally, particles have hypothetical ball states that include position, velocity and acceleration in a 3D coordinate system.

PF generally makes use of appearance features to identify the ball. Selection of appearance feature usually depends on the type of ball to be tracked. Tracking of balls in indoor sports like Snooker (Rea et al. 2004), Billiards, Tennis and Pool typically use color as a strong feature in PF. Conventional PF has also been applied for locally searching a ball, lost due to small size or occlusion, in a switching search method proposed by Arika et al. (2008). Per-frame ball tracking is implemented in ASPOGAMO (Beetz et al. 2009) where current state of the ball is updated as per the ball dynamics. However, tracking in presence of motion blur and occlusion remain partially solved even using PF. Performance of the ball tracking algorithm can be improved using accumulation image along with PF, after removing player and noise blobs (Choi and Seo 2004, 2005). The accumulation image not only provides a proposal density for PF, but also decides whether the ball was present in the image frame or not.

In order to cope with the cluttered background, occlusion and motion blur, a few modifications have been incorporated in the conventional PF. An adaptive PF is designed by Huang et al. (2008) where a strong detector is used to trigger the tracker while a weak detector is used to enhance the tracker. Strong detector is generated from shape properties of foreground blobs. However, the weak detector is generated using the foreground detection likelihood and

Table 4 Summary of Kalman filter based approaches

References	Sport	Tracking methodology	Dataset	Result	Strength	Shortcomings
Ishii et al. (2007)	Soccer	KF	2 cameras captured videos at Kasumigaoka stadium, Japan	Qualitative	Can detect ball bounce	Cluttered background, occlusion
Ren et al. (2006)	Soccer	KF	3 seq of 4800 frames each	Tracked rate of 46.93, 75, 78.36, 78.67 and 78.86% for buffer sizes of 0, 25, 50, 75 and 100 frames	No false alarms	Inaccurate modeling of ball
Chen and Zhang (2006)	Table Tennis	KF + Incremental Bayesian algorithm	2004 Olympic Female Final, 2000 Olympic Male Final and 2005 World Cup Male Final, 352 × 288 at 25 fps	TP—95.62%, TR—74.92%	Both dynamic and appearance information of the ball are used	Low-resolution video, occlusion
Ren et al. (2009)	Soccer	KF	8-camera system, 4800 frames, 720 × 576, 25 fps	TA of 94.1, 91, 89.2 and 88.5% for buffer sizes of 0, 25, 50 and 75 frames	No false alarms	Worked only on long shots
Kim and Kim (2009)	Soccer	DKF	2008–2009 English Premier League Round 8 Chelsea versus Middleborough, 592 × 320, 1000 frames	Average error in pixels is 20.73 and DTA of 66.8% for a threshold of 10 pixels	Robust ball tracking, partial occlusion handled	Fails when players are crowded, High time complexity
Zaveri et al. (2004)	Ping pong, Cricket	KF + NNDA	100 frames ping pong, 32 frames of cricket	Quantitative	Handles noisy background and changing ball size, fast method	Too small dataset, occlusion

Table 4 continued

References	Sport	Tracking methodology	Dataset	Result	Strength	Shortcomings
Desai et al. (2005)	Ping pong, Cricket	KF + NNDA	100 frames ping pong, 32 frames of cricket	Qualitative	Handles small, fast, low contrast, negligible textured balls, implemented on Xilinx Virtex-II Pro FPGA	Too small dataset, occlusion
Kittler et al. (2005)	Tennis	KF + DA	Approx. 30 min play from Woman's Final of 2003 Australian Open	No Tracking results	Ball bounce or hit is modeled in multiple hypothesis	Time consuming during long-term occlusion
Liang et al. (2005)	Soccer	KF + TM	Several video clips with 200 to 1000 frames, 352 × 288, 25 fps	TP—89.1%, TR—84.5%	Fast tracking method, works even in bad playfield conditions	Occlusion, misdetection, can detect only circular objects
Liu et al. (2006)	Soccer	KF + TM	2 video sequences with 600 frames each	TA—94.77%, TP—89.1%, TR—84.5%	Can estimate ball's height with no reference	Occlusion, size, shape changes due to fast motion not considered
Liang et al. (2007)	Soccer	KF + TM	2 video clips with a total of 1369 frames	TP—89.4%, TR—80.7%	Can estimate ball's height with no reference	Long-term occlusion, size and shape changes

seq sequence, *TM* template matching, *DKF* dynamic Kalman filter, *DA* data association, *NNDA* nearest neighbor data association, *TA* tracking accuracy, *TP* tracking precision, *TR* tracking recall, *DTA* detection accuracy

gets integrated into observation likelihood of the tracker. The mixture model of the two detectors overcomes the ambiguity of template matching (TM) in cluttered background. Tracking becomes complicated when any model is learning the motion of a ball that has non-linear and non-periodic oscillatory movements like Table Tennis ball. To overcome the problem, [El Abed et al. \(2006\)](#) incorporated an adaptive dynamical model into PF. In order to adapt to the ball's size, color and texture changes, [Tong et al. \(2004\)](#) appended Factored Sampling based Condensation algorithm with region optimization for soccer ball tracking. The authors also defined a scale-invariant soccer ball template confidence measure in order to bootstrap the re-detection of the ball region needed for continuous tracking. Though the system yields real-time tracking results, it again failed under occlusion, complex background and small ball size conditions. [Yan et al. \(2005\)](#) used two separate models for tracking tennis ball either in flight mode (first dynamic model) or in transition mode (second dynamic model). This improved sampling efficiency based PF algorithm has two advantages viz.; The Monte Carlo simulation limits the computational complexity by choosing the number of particles without losing Bayesian optimality. Another advantage is that the samples are directly drawn from the posterior density instead of drawing from the prior probability distribution. A Support Vector Regression (SVR) based PF (SVRPF) reweighting scheme is introduced by [Zhu et al. \(2009\)](#) in order to re-approximate the posterior density in PF. Sample distribution in SVRPF is maintained to achieve detection after out of view situations and momentary occlusions.

[Misu et al. \(2007\)](#) proposed an architecture for soccer ball tracking using distributed PF. The filter exchanges particles among tracking units. The system improved positional accuracy along with robustness against factors like; occlusion, ball out of frame, system instability and time-out situations due to hanging. This could be possible by integrating multi-view information from two cameras. None of the above approaches works on long video sequences. [Cheng et al. \(2015\)](#) explored the likelihood estimate of PF and incorporated ball size adaptive (BSA) tracking window. The corresponding ball feature likelihood model and an anti-occlusion likelihood measurement model is proposed for ball tracking in 3D in volleyball sports using 4 cameras. However, it is very difficult for the system to recover the tracking targets once lost. [Cheng et al. \(2016\)](#) removed the shortcoming by incorporating automatic recovery of a lost target by 3D detection and anti-occlusion observation model. However, deleting some camera information with low likelihood makes the particles unstable. Both the approaches given above are not suitable for low-resolution videos with less than standard framerate (30 fps). This is because in low-resolution videos, fast moving ball appears blurred, elliptical, but not circular as expected in the above two approaches.

PF based approaches work by tracking the particles distributed all over the target. A few methods claim to satisfactorily overcome problems like small sizes, varying scales, partial occlusion, back-ground clutter, non-linear and non-periodic motion. However, when size of the ball becomes very small or when its color appears different from original then the number of particles reduces drastically which leads to failure of the tracker. When the ball appears touching a player wearing similar color uniform or when the ball merges with similar appearing color background, the particles assigned to the ball are distributed over both the regions. In such cases, if good redetection mechanism is not incorporated in the tracker, then it may lead to permanent failure of the tracker. In addition, the particles are not rapidly updated which leads to poor performance in tracking fast moving balls. One solution may be to reduce the number of particles assigned to a ball but it may result in increased number of false positives. This approach can work for close-up and medium shots, but gives poor performance for long shots due to small size of the ball. Table 5 presents a summary of all PF based ball-tracking methods published so far in different sports, validating databases, obtained results, strengths and shortcomings.

Table 5 Summary of particle filter based approach for ball tracking

References	Sport	Tracking methodology	Dataset	Result	Strength	Shortcomings
Rea et al. (2004)	Snooker	Color based PF	Broadcast Snooker footage, 30 min	NA	Simple	Specific to cue sports like Snooker, Pool and Billiards
Ariki et al. (2008)	Soccer	PF	38th National high school championship finals Kyoto, Japan, 800 frames, 1280 × 720 pixels	TRT—91.8%, Processing time—0.9 fps	Handled ball lost situation	Size and shape changes, occlusion
Zhu et al. (2008)	Soccer	PF	FIFA World Cup 2006, 704 × 576	D and T—88.6%	Trained Adaboost classifier is used for ball trajectory identification	Occlusion, false positives in detection
Beetz et al. (2009)	Soccer	PF	World Championship Finals 2006	Qualitative	Motion blur, partial occlusion handled	High time complexity
Choi and Seo (2004)	Soccer	PF	8000 frames stitched, 1361 × 480	Qualitative	Handled full occlusion, no manual intervention	Size, shape changes not considered, high time complexity
Choi and Seo (2005)	Soccer	PF	600 frames, 960 × 540	Qualitative	Accumulation image increased the performance, handled full occlusion, no manual intervention	Size, shape changes not considered, high time complexity
Huang et al. (2008)	Soccer	Adaptive PF	2 test videos, France, 260 frames, 360 × 288 and Belgium, 375 frames, 360 × 240	Qualitative	Handles occlusion, heavy clutter and motion blur, small object detection.	Very small database, complex background
El Abed et al. (2006)	Table Tennis	Adaptive dynamic model based PF	90 frames	Qualitative	Can track non-linear, non-periodic oscillatory moving objects	Very small database, specific to TT, occlusion, size, shape changes

Table 5 continued

References	Sport	Tracking methodology	Dataset	Result	Strength	Shortcomings
Tong et al. (2004)	Soccer	Factored sampling based Condensation filter	Sequence 1–50 frames, Sequence 2–170 frames	Ball D, T and region optimization—1218 ms for seq 1, 2199 ms for seq2	Redetection of object for continuous tracking	Occlusion, complex background and small ball size. Very small dataset
Yan et al. (2005)	Tennis	PF with improved sampling efficiency	2003 Australian Open tennis tournament,	Qualitative	Detect elliptically appearing ball	Complex background, redetection not possible
Zhu et al. (2009)	Soccer	Support vector regression PF + TM	FIFA World Cup 2006, 704 × 576	Average P and R of 88.4 and 85.1% respectively	Momentary occlusion and ball going temporarily out of view	Occlusion, misdetection
Misu et al. (2007)	Soccer	Distributed PF	—	—	Robust to occlusion and system instability	Only long shots considered
Cheng et al. (2015)	Volleyball	PF	3 sets of Inter High school Games Men's Volleyball, Tokyo 2014, 1920 × 1080, 60 fps	SR of 3D ball T—93.39%	Changes in ball size, complex background, partial occlusion	High time complexity, redetection not possible
Cheng et al. (2016)	Volleyball	PF	3 sets of Inter High school Games Men's Volleyball, Tokyo 2014, 1920 × 1080, 60 fps	SR of 3D ball T—99.14%	Automatic recovery handles failure during redetection	Removing some camera information with less likelihood disturbs the position of particles, not suitable for low-resolution videos

seq sequence, TM template matching, D detection, T tracking, TRT tracking rate, R recall, SR success rate

4.3 Trajectory

Trajectory based ball tracking approach is a wide umbrella covering tracking techniques that use physical or dynamic models, physics-based approaches, TM and triangulation. Since, major focus of these methods is tracking a ball using its trajectory; they are grouped as a single approach. Major focus of various researches done under this approach is building an optimum 2D or 3D ball trajectory that is a kind of representation for tracking a ball on sports ground. Generally, it is an offline procedure where ball candidates are obtained by detecting them over a sequence of frames. They are joined together using some prediction measures, after filtering false positives and missed observations; finally, a true ball trajectory is obtained. According to this approach, detecting a ball among ball candidates is difficult than detecting whether a trajectory is the ball trajectory or not.

Initial efforts are more focused on obtaining 3D ball positions using basic TM method. They use ball region from current frame as template for the next frame, search in a limited region around the previous position and use back tracking to confirm the result (Ohno et al. 1999; Yow et al. 1995). If a ground model is available, start and end-points of the ball trajectory are located on the ground. If a perpendicular object is located on the ground, then triangular geometric relations can be used to determine the 3D ball position (Kim et al. 1998). If the end-points are unavailable then a physics based search is used to determine them. Reid and North (1998) improved over Kim et al. (1998) using ball's shadow on the ground as a reference for predicting 3D ball position thereby reducing the dependability on physics based ball model. However, selection of perpendicular object and shadow was a manual process. A few approaches (Ren et al. 2004a, 2008) categorized and modeled the ball motion in 4 phases' viz., rolling, flying, in-possession and out-of-play. They did not require additional cues like reference player (Kim et al. 1998), goalmouth (Kim et al. 1998) and shadows (Reid and North 1998) for 3D position estimation of a ball in soccer. A few researchers (Ohno et al. 2000; Yamada et al. 2002) estimated ball positions in 3D space by fitting a physical model of a ball movement to the observed ball trajectory. However, human intervention is needed to guess the ball speed from image and that is a major limitation of this approach.

In order to deal with cases of occlusion, Seo et al. (1997) employed backprojection technique but required manual input of starting locations. A multi-threaded approach was used for tracking Tennis ball using multiple camera setting (Pingali et al. 2000, 2002). Depending on current ball velocity, the system is continuously updated. Segmentation and tracking parameters like intensity value ranges, expected size, search window size and different thresholds are dynamically updated. However, in order to speed up the process, monochrome cameras were preferred thereby sacrificing a strong cue i.e. color of the tennis ball. After ball detection, Liu et al. (2006) applied adaptive TM approach and its hybrid with KF for tracking. It can be observed that most of the template-based methods require either manual assistance to perform key tasks like selection of perpendicular object/shadow or frequent update of template. This reduced the efficiency, introduced errors and cannot be considered automatic. TM fails for detecting fast motion of the ball as its appearance changes with change in velocity. Even adaptive TM was not able to continue tracking.

In subsequent works, ball detection processes became more automated due to the use of background subtraction algorithms for obtaining moving regions in the form of blobs. The blobs are processed and several filters (size, shape, color criteria etc.) are applied on them to obtain ball candidates. Chen et al. (2007, 2012) classified the ball candidates into isolated and contacted candidates. In order to ease the extraction of ball trajectory from large number of noisy and false candidates, 2D plots (Candidate Feature Image (Yu et al. 2003a), X-Distribution Image and Y-Distribution Image (Chen and Wang 2007), X Time Plot and Y

Time Plot (Chen et al. 2012), Candidate Distribution Plot (Chakraborty and Meher 2012a)) or 3D plots (Miura et al. 2009) are generated by plotting the centroid of ball candidates. For example, X-Time plot is the plot of input video frame number on X-axis and x-coordinate of the ball position in the frame on Y-axis. Initially, all the ball candidates are linked to their nearest neighbor candidate in the next frame forming large number of trajectories. Depending on the observed ball movement along Y and X-axes, estimation functions are modeled; one parabolic function for a ball trajectory along Y-axis and another straight line function for projection of the trajectory along X-axis. These estimation functions are used to predict ball positions in the subsequent frames (Chakraborty and Meher 2012a, b; Chen et al. 2007, 2009, 2012). For non-team sport like Tennis, a ball track is formed by blobs following a linear trajectory (Conaire et al. 2009). Trajectory is grown by adding ball candidates that are present at estimated positions and the estimation function is updated using regression. A few methods replaced the estimation function either with velocity constraint (Chen et al. 2009) or KF based candidate verification (Yu et al. 2003a, b, c, 2004b, 2006, 2007b) method or KF verification function (Chakraborty and Meher 2012b). Instead of single 2D KF function, Ekinci and Gokmen (2008) and Yu et al. (2004b) used two 1D KF for eliminating the need of two separate dynamic models as well as information of player positions. Next stage deals with identification of ball trajectory using either least square fitting technique (Chen et al. 2009) or assigning confidence degrees to each trajectory (Chen et al. 2007, 2012) or confidence index (Yu et al. 2003a, c, 2006, 2007b) to the trajectories. The best fitting trajectory or one with highest confidence degree is chosen as the ball trajectory. The obtained ball trajectory is extended by interpolating and patching the gaps resulting into actual ball trajectory. If the gap is large then localized KF based model matching is used to narrow down the gap (Yu et al. 2003c). However, lack of good initial detection and subsequent intermediate detection methods leads to failure in object segmentation resulting into improper trajectory.

Obtained 2D ball trajectory, calibrated camera parameters from multiple sources and physical characteristics of ball motion are used to reconstruct 3D trajectory of the ball (Chen et al. 2009). The obtained ball locations can be superimposed on original video frame for better visualization of ball trajectory (Chakraborty and Meher 2012a). For detecting position of the ball in a 3D trajectory, Ren et al. (2004a, 2008) used two methods. First one was triangulation using visual information from multiple sources. The other method was to analyze trajectory from a data source and infer that the ball is in flight. The ball trajectory was approximated by straight lines on a virtual playfield (Leo et al. 2008). Kumar et al. (2011) also used triangulation of 2D trajectories obtained from two cameras for obtaining 3D trajectory of the ball. They used basic kinematic equations to predict locations of missing balls in 3D space. However, most of the techniques discussed above did not take bouncing phenomenon of a ball into account. Hawk-eye Innovations Ltd (2015) work for ball tracking in sports is based on the concept of triangulation from multiple high-speed HD cameras. Each camera determines the ball position and plot the ball trajectory in 2D. Central system combines all the 2D trajectories to form a 3D trajectory and visualize any shot in 360°. Under the availability of initial conditions and environmental parameters, Poliakov et al. (2010) used physical model (a flight model and a bounce model) for accurately simulating ball trajectory. They used genetic algorithm to fit the observations with the candidate trajectories. They managed to do ball tracking and 3D reconstruction of trajectories simultaneously with good resistance to noise and outliers.

A very important advantage of this approach is that they can deal with missing, merged or partially occluded ball in a few frames. Most of the methods need not remove salient objects like lines on the ground and basketball board for successful ball tracking. However, most of the researches considered long shots with full trajectories, i.e. a situation easier to work upon,

compared to the close-up and medium shots with a lot of dynamic activity and shorter passes. Change of appearance in size, shape and color of ball due to high velocity and illumination changes are not considered. Most of the methods use KF for position estimation; however, it does not work for segments containing fast and curved ball movements. In presence of multiple moving ball-like objects in the background, it may lead to failure of the algorithm. Calculation of camera parameters for 2D to 3D conversion of ball trajectory is a computationally complex task. Ignoring the effect of air friction, complex ball movements like swerve, ball spin axis, ball spin rate, etc. may also lead to inefficient ball tracking. Commercial ball tracking solutions have a strong limitation of viewing angle, high cost of camera system and its installation, visible area constraint and are impractical for research in general. Table 6 presents a summary of all trajectory-based ball tracking methods published so far in different sports, validating databases, obtained results, strengths and shortcomings.

4.4 Data association

Data Association (DA) is a process of associating uncertain measurements (ball position in image obtained after detection) to known tracks. It requires detected ball candidates as input. The difficulty level of DA increases with increase in number of targets to be tracked. DA problem in ball tracking uses a Markov Chain assumption and a predictive motion model. However, in presence of noisy detections, the problem of false positives and false negatives makes ball tracking using DA, a computationally complex process in complex environments.

DA has been used in object tracking since its introduction almost three decades ago (Bar-Shalom 1987). However, it was used in ball tracking for the first time in the form of Robust Data Association (RDA) (Lepetit et al. 2003). RDA treated DA as a motion-fitting problem. They use RANSAC-like algorithm to solve the DA and estimation problem as it is robust to abrupt motion changes. However, as the ratio of true positives drops, RANSAC becomes computationally more complex in a highly cluttered environment. In RDA, multiple models give periodical estimates that are independent of each other. Since, global smoothness constraint is not applied, wrong pick of a clutter-originated motion as the best model may result in an irreparable error. Yan et al. (2006a) used Viterbi algorithm to search the sequence of measurements with a posteriori probability. They used bidirectional Viterbi search method to improve ball tracking by recovering from errors caused by abrupt change in ball motion. Yan et al. (2008, 2014) proposed a hierarchical layered DA scheme with graph theoretic formulation to overcome the shortcomings of RDA. In the first pass, the framework uses a recursive estimation scheme to fit a dynamic model. In the second pass, Dijkstra's shortest path algorithm filters out the false tracklets to find the best trajectory. While Dijkstra's algorithm is a single-source shortest path algorithm, Yan's algorithm requires manual labelling of initial and final tracklets. The initiation and termination problems were dealt by studying all-pairs of shortest paths obtained from single object tracking (Yan et al. 2008). They did not provide any mechanism to handle cases of a ball going out of field and frequently occurring short-term misdetections.

Huang et al. (2012) proposed a Viterbi algorithm to estimate a globally smooth ball trajectory. However, they incurred some loss of precision in the obtained trajectory due to absence of a well-defined kinematics motion model. Ball tracking has also been done either using nearest neighbor DA (Zaveri et al. 2004) or DA with KF (Kittler et al. 2005). Zhou et al. (2015) proposed a two-layered DA for tennis ball tracking which is an extended version of Zhou et al. (2013). The proposed approach uses both; the global and local information to search the possible trajectories efficiently. The authors claim that the proposed approach for single object tracking is extendable for tracking multiple objects as well. Unlike Huang

Table 6 Summary of trajectory based approach for ball tracking

References	Sports	Tracking methodology	Dataset	Result	Strength	Shortcomings
Yow et al. (1995)	Soccer	TM + Physics based	World Cup 1994	NA	Handled scale changes	Occlusion, high speed, low detection rate
Kim et al. (1998)	Soccer	Trajectory + Physics based	NA	NA	Simple	Only long shots, occlusion, ball does not move in a parabolic motion in actual case
Reid and North (1998)	Soccer	Trajectory + TM + KF	Euro 1996 match	Qualitative	Ball's shadow used as reference	Occlusion, high speed, only long shots, reference dependent, difficult to detect shadow
Ohno et al. (1999)	Soccer	Physics + TM	320 × 240, 30 fps, 300 frames	Qualitative	Player position used as reference	Full occlusion, high ball speed, size and shape changes, ball tracking in easy cases, poor performance for BSV
Ohno et al. (2000)	Soccer	Physics + Trajectory + TM	NA	Qualitative	3D position estimation of ball	Manual operator needed, inaccurate modeling of 3D ball trajectory, high ball speed, size and shape changes
Pingali et al. (2000)	Tennis	Multi-threaded tracking, TM	NA	Qualitative	Real-time, scalable, dynamic parameter update	Ball occlusion due to players, tracking ball only during serve
Pingali et al. (2002)	Tennis	Multi-threaded tracking, TM	NA	Qualitative	Real-time, scalable, dynamic parameter update	Ball occlusion due to players, tracking ball only during serve
Yamada et al. (2002)	Soccer	Physics + Trajectory + TM	NA	Qualitative	3D position estimation of ball	Occlusion, illumination changes, high ball speed, size and shape changes

Table 6 continued

References	Sports	Tracking methodology	Dataset	Result	Strength	Shortcomings
Yu et al. (2003a)	Soccer	Trajectory + KF	1000 frame sequence from 2002 FIFA World cup	A 1st seq—86.4%, A 2nd seq 77.9%	Better ball identification	Poor result in close-up shots
Seo et al. (1997)	Soccer	Trajectory + KF + TM	NA	Qualitative	Occlusion	No ball detection method, need manual assistance
Yu et al. (2003c)	Soccer	Trajectory + KF	1000 frame sequence from 2002 FIFA World cup	A 1st seq—100%, A 2nd seq—93.3%	Focus on finding ball trajectory than finding ball region.	Poor result in close-up shots, high ball speed
Yu et al. (2003b)	Soccer	Trajectory + KF + TM	15 seq	A—96%	Size changes handled	Shape and illumination changes
Yu et al. (2004b)	Tennis	Trajectory + KF + TM	7 seq of total 120 s, Men's final of French Open 2003	Average A—98.94%	Used domain knowledge of Tennis, better ball localization	Poor ball segmentation
Ren et al. (2004a)	Soccer	Triangulation + Trajectory	5500 frames of multi-cam seq.	Qualitative	No reference needed for 3D ball position estimation, phase specific trajectory estimation	Occlusion, high ball speed
Yu et al. (2006)	Soccer	Trajectory + KF	2 videos from 2002 FIFA World cup	A—81%	Detect partially or fully merged ball	Computational time not reported, time complexity seems on higher side
Liu et al. (2006)	Soccer	KF + adaptive TM	2 videos, 600 frames each, 720 × 572	DP—93.5%, R—58.75%. (D + T)P—89.4%, (D + T)R—80.6%	No need of Human intervention	Occlusion, high ball speed

Table 6 continued

References	Sports	Tracking methodology	Dataset	Result	Strength	Shortcomings
Chen et al. (2007)	Volleyball	Trajectory + Physics based	Set1—Asia men's Volleyball Challenge cup, Taiwan, Set2—10 clips from internet, 352 × 240	D—71.28%, T—91.68%	Proposed curve fitting algorithm for volleyball seq, generic for games with similar ball motion characteristics	High ball speed, tracking fails if detection is poor
Chen et al. (2008)	Baseball	Trajectory	Major League Baseball, Japan Professional Baseball, Chinese Professional Baseball League	D—92.43%, T—96.62%	Simple	Cluttered background
Chen et al. (2009)	Basketball	Physics + Dynamic Programming + Trajectories	BBV, Olympic US versus Spain, Olympic US versus China, Taiwan high school basketball league, Korea basketball game	R—92.83%, P—95.41%	Better method of 2D to 3D inference, generic trajectory evaluation approach	High computational complexity as it calculates camera parameters for 2D to 3D conversion, false alarms
Ekinci and Gokmen (2008)	Tennis	Trajectory + KF	5 videos from Dubai 2008	SR—94.59%	No need of separate dynamic models and knowledge of players positions	False positives, high ball speed
Ren et al. (2008)	Soccer	Trajectory + Triangulation	Fulham Football club, UK, 2001, more than 5500 frames, 720 × 576	D—85%	No manual assistance needed	Severe occlusion, insufficient observations, ignores air friction and complex ball movements like swerve
Leo et al. (2008)	Soccer	Trajectory + Triangulation	Italian soccer championship 06–07, 1920 × 1080, 25 fps	Qualitative	Good for event detection	Occlusion, high ball speed
Conaire et al. (2009)	Tennis	Trajectory	Training match	Qualitative	Ball tracks used to infer serves and rallies	Misdetections, only long shots

Table 6 continued

References	Sports	Tracking methodology	Dataset	Result	Strength	Shortcomings
Poliakov et al. (2010)	Tennis	Physics + Trajectory	30 min video footage	Qualitative	Robust to noise, flexible, used ball spin	High ball speed, occlusion
Kumar et al. (2011)	Tennis	Trajectory + Triangulation	NA	Qualitative	Simple	Air resistance, ball spin neglected, occlusion, high speed
Chen et al. (2012)	Volleyball	Trajectory	352 × 240, 30 fps, Asia Men's Volleyball Cup 2005	DA—71.85%, TA—87.12%	Multi-stage approach for ball detection and trajectory identification	Poor object segmentation, air friction, ball spin axis, ball spin rate not considered, 2D–3D conversion consumes time
Chakraborty and Meher (2012a)	Basketball	Trajectory	Video of 640 × 480, 30 fps	DA—94.77%, TA—100%	Handled occlusion and merged balls, simple	Compared with simple KF based methods and not on other trajectory based methods, high ball speed, small pass not considered
Chakraborty and Meher (2012b)	Volleyball	Trajectory + KF	6 BVV	DA—91.8%, TA—96.76%	Almost real-time	High ball speed, small pass not considered,
Chakraborty and Meher (2013b)	Basketball	Trajectory + KF	6 BBV	DA—87.06%, (D + TA)—93.18%	Handled dynamic background, different illumination conditions	Only long shots are considered, high speed, air friction, ball spin not considered

Table 6 continued

References	Sports	Tracking methodology	Dataset	Result	Strength	Shortcomings
Chakraborty and Meher (2013a)	Basketball	Trajectory	BS seq—344 frames, DS seq—254 frames	BS seq—DA—96.22%, TP—100%, DS seq—DA—92.52%, TP—100%	Fast, robust to color changes, occluded and merged balls. Implementation is simple and easy.	Ball spin, air friction, small video sequences, high ball speed, long shots only.

seq sequence, *A* accuracy, *D* detection, *T* tracking, *R* recall, *P* precision, *DP* detection precision, *DA* detection accuracy, *TA* tracking accuracy, *SR* success rate, *BVV* broadcast volleyball videos, *TM* template matching, *BSV* broadcast soccer videos

et al. (2012) and Zhou et al. (2015) use a motion model and a few measuring scores for acquiring precise trajectories locally using shift token transfer algorithm. At the local layer, the authors introduced virtual ball candidates for handling long-term absences and short-term misdetections. At global layer, they used Floyd algorithm that searches trajectories over a weighted directed acyclic graph and thus avoid manual labelling of initial and terminal tracklets. They introduced K-order directed arcs over the graph to tackle the problem of absence of ball in long video sequences.

DA algorithms are prone to misdetections, noisy background and cannot track ball temporarily going out of field. DA has been applied for ball tracking only in Tennis sport where mere two players are present on the court, occlusion is very less and field-of-view is limited. In addition, DA is an offline method of ball tracking. There is a good scope for applying DA to track ball in team sports like Volleyball, Soccer, Basketball, etc. It may require some modifications in the present hierarchical approaches. Table 7 presents a summary of all DA based ball-tracking methods published so far in different sports, validating databases, obtained results, strengths and shortcomings.

4.5 Graph

Graph based approach for ball tracking involves finding actual ball route from a graph constructed out of detected ball candidates and representing the required trajectory. Usually features like color, size, shape, motion, form factor, etc. are used for finding ball candidates in each frame. Pallavi et al. (2008) deliberately avoided the use of color feature as it is prone to changes in illumination and velocity and still presented a robust framework to poor playfield conditions without any pre-processing. Instead, they use circle-detection method called circle Hough transform (CHT). When almost all the methods focus on long shots for tracking ball, Pallavi et al. (2008) focused on medium shots as well. They filtered the non-ball candidates in medium shots using motion information and in long shots using size, position, velocity and direction of motion. A few methods formed transition graph to represent ball transitions between objects like players, ground lines, stands, and other ball candidates (Miura et al. 2009; Shimawaki et al. 2006a, b). After obtaining the ball route candidates using spatiotemporal relations, the authors searched each ball candidate with a ‘Separability filter’, clustered ball-like regions and formed cluster segments by fitting a line to each cluster. Other methods used ball candidates to form a weighted graph where each node represents a ball candidate and link (edge) representing possible transitions between nodes. Each node is assigned a weight that denotes its resemblance with the ball. Each edge is assigned a weight that represents correlation between the joined candidates (Liang et al. 2005; Pallavi et al. 2008; Yu et al. 2007a). Ball route candidates are estimated by applying dynamic programming [Viterbi algorithm (Liang et al. 2005; Yu et al. 2007a)] on the graph using some constraints [velocity (Shum and Komura 2004)]. In order to finely tune the estimated route, Shum and Komura (2004) used B-spline curve for interpolating the missing nodes. For finding the actual ball trajectory, a few researchers (Miura et al. 2009; Shimawaki et al. 2006a, b) selected a ball candidate trajectory with the highest allotted score whereas Yu et al. (2007a) used least square method. Researchers Shum and Komura (2004) constructed 3D trajectory of a ball using its 2D trajectory and an aerodynamic model.

Similar to DA algorithms, graph based approach can also handle occlusion to some extent. However, if the ball is missing for a large number of frames then the computational complexity of the algorithm increases and may lead to permanent discontinuity between candidate trajectories. A few methods were unable to detect ball moving out of field. None of the methods in this approach discussed the cases of size and shape changes of ball due to high velocity and

Table 7 Summary of data association based approach for ball tracking

References	Sport	Tracking methodology	Dataset	Result	Strength	Shortcomings
Zhou et al. (2015)	Tennis	Two-layered DA	Australian Open 2010, 1280 × 720, 37,206, 25 fps. U.S. Open 2011, 1280 × 720, 24,363, 25 fps	P, R and F1 of 84.39, 75.81 and 79.87% resp. for Aus Open. 82.34, 67.01 and 73.89% resp. for US Open	Good balance of precision and recall	False positives
Huang et al. (2012)	Tennis	Viterbi algorithm	Wimbledon 2008, 180 min for Training. Australian Open 2010, 106 min. US Open 2011, 81 min. French Open 2012, 83 min for Testing	Qualitative	Combining audio with video gives better performance than using video alone.	Fails for cluttered background and imprecise camera calibration
Huang et al. (2011)	Tennis	Graph theoretic DA	Four video clips of women's double Australian Open 2008 for training and testing with 37.07, 33.48, 37.56 and 13.40 min resp. Audio information was used from Men's single Wimbledon Open 2008	NA	Combining audio with video gives better performance than using video alone.	Clutter background
Zhou et al. (2013)	Tennis	Two-layered DA	Australian Open 2010, 1280 × 720, 37,206, 25 fps. U.S. Open 2011, 1280 × 720, 24,363, 25 fps	P, R and F1 of 73.13, 73.11 and 73.12% resp. for Aus Open. 82.31, 67.52 and 74.19% resp. for US Open	Robust to false detections	False positives, high ball speed, not applicable to small size views

Table 7 continued

References	Sport	Tracking methodology	Dataset	Result	Strength	Shortcomings
Yan et al. (2014)	Tennis	Layered DA	Women's Australian Open 2003, 25 sequences, 15716 frames for Training. Men's Australian Open 2006, 70 sequences, 60094 frames and Men's Wimbledon Open 2005, 70 sequences, 43706 for Testing.	Average F of 0.7435	Handle complexities in BTV. Can remove false tracklets. Insensitive to small changes of parameters.	Method is sensitive to increase of tracklets threshold. If it increases, ball tracklets are discarded
Yan et al. (2008)	Tennis	Layered DA	Women's Australian Open 2003, 25 sequences, 15716 frames for Training. Men's Australian Open 2006, 70 sequences, 60094 frames and Men's Wimbledon Open 2005, 70 sequences, 43706 for Testing.	Average F of 0.7435	Handle false candidates very quickly, efficient	Misdetection, high ball speed
Yan et al. (2006b)	Tennis	Layered DA	Women's final Australian Open 2003, 25 sequences, 5 min long	Qualitative	Handle low quality video, abrupt model changes and heavily cluttered environment.	Misdetection, high ball speed

Table 7 continued

References	Sport	Tracking method-ology	Dataset	Result	Strength	Shortcomings
Yan et al. (2006a)	Tennis	Bidirectional maximum a posteriori probability-Viterbi DA (MAP-VDA)	Women's final Australian Open 2003, 25 sequences, 5 min long	Qualitative	Can handle abrupt ball motion changes.	Computational complexity increases as the missing candidates increase.
Lepetit et al. (2003)	Tennis	Robust DA	About 1000 de-interlaced frames	Qualitative	Treated DA as a motion fitting problem	Background clutter, size changes. Cannot detect ball during bouncing.

P precision, R recall, $F1/FI$ measure, F F-measure, A_{tas} Australian

heavy occlusion. Table 8 presents a summary of all graph-based ball-tracking methods published so far in different sports, detection techniques, validating databases, obtained results, strengths and shortcomings.

4.6 Optimizing an objective energy function

Generally, in team sports, ball-tracking algorithms discussed in previous approaches face difficulty due to frequent occlusion or possession by players, unpredictable ball and player movements during short passes. Ball tracking during long passes is comparatively easy owing to the fact that the ball follows second order motion model. In addition, the previous approaches were confined to a sport without any scope for generalization. These approaches track ball and players separately thereby requiring two different methods that unnecessarily increased the computational complexity of the whole framework.

In order to overcome the problems discussed above, Wang et al. (2014a) formulated the ball tracking method by deciding which player is in possession of the ball at a given time. They introduced a focus on scene (FoS) tracker which is based on Conditional Random Fields (CRF) model. The tracker minimizes a global objective function that takes into account player trajectories, phase of the game and image evidence (detection) of the ball. The system showed consistent high performance when applied to long basketball and soccer video sequences. It achieved a processing speed of 8–10 fps with good results but with misdetection in a few frames. Wei et al. (2014) extended the work in Wang et al. (2014a) by finding location of a ball in short-term future instead of finding the ball captivator at the present instance. Wei et al. (2014) used augmented-hidden conditional random field (a-HCRF) which incorporates local observations in hidden conditional random field (HCRF), boosting its prediction performance. In order to reduce dependency in tracking, Wang et al. (2014b, 2016) tracked both the ball and players simultaneously using intertwined flow variables. The authors could estimate the trajectories of the ball, which were not detected initially based only on image evidence, by estimating the trajectories of players and ball jointly. Authors formulated the problem of simultaneous tracking as a constrained Bayesian inference problem where one of the players' possesses the ball. They formed a directed acyclic graph (DAG) of the discretized locations and poses that were reframed into mixed integer program (MIP). They minimized the objective function subject to the flow constraints. The authors reduced the computation complexity initially by pruning the graph and then by introducing tracklet based formation of the optimization problem. Though this method is generic and applicable to many team sports, it cannot preserve the identity of an object during momentary occlusion and performs misdetection a few times. Their ball possession is modeled explicitly and they perform tracking on discretized grid with no physical constraints that reduces smoothness of the trajectory. Maksai et al. (2015) explicitly model ball and player interactions and physical constraints obeyed by ball during flight. Initially the ball-tracking problem is formulated as maximization of a posteriori probability using factor graph that was reformulated in terms of an integer program using Ball graph. Further, adding various constraints, the final problem was addressed using MIP. Similar to Wang et al. (2016) and Maksai et al. (2015) introduced a ball tracking method by taking ball-player interactions into consideration. Additionally, they imposed physics based constraints. Their method described ball behavior in a better manner than other approaches.

The energy optimization approaches formulated the ball-tracking problem in a more precise manner compared to other approaches. They considered various phases of ball viz., in flight, in possession, occlusion and applied suitable constraints to the objective function thereby taking fine details of the ball state into consideration. They yield superior performance

Table 8 Summary of Graph based approach for ball tracking

References	Sports	Ball detection	Ball trajectory generation	Dataset	Results	Strength	Shortcomings
Shum and Komura (2004)	Baseball	Color, shape, motion and proximity with ball candidates	Graph method, Dynamic programming, B-spline curve	NA	Qualitative	White-on-white scenario taken care of	High ball speed and change in color due to speed
Shimawaki et al. (2006b)	Soccer	Transition graph	Spatio-temporal relations, separability filter; Score is used to identify the actual ball route	17,279 frames	Qualitative	Fine tuning the detections	Does not tell whether the occluding player has touched the ball or not. Out of field balls
Shimawaki et al. (2006a)	Soccer	Color, size, shape wise ball detection. Ball tracking using simple prediction,	Spatio-temporal relations, separability filter; Score is used to identify the actual ball route	17279 frames	Qualitative	Fine tuning the detections	Does not tell whether the occluding player has touched the ball or not. Out of field balls
Miura et al. (2009)	Soccer	Color, size, shape wise ball detection. Ball tracking using simple prediction,	Spatio-temporal relations, separability filter; Score is used to identify the actual ball route	First video 480 × 240 second video with 17279 frames	Qualitative	Occlusion by players, redetection	High ball speed, size, shape changes
Yu et al. (2007a)	Soccer	Size, form factor, ratio of object area to bounding box	Weighted graph, Viterbi based algorithm, Least square method	2 video clips with more than 500 frames FIFA World Cup 2006	Qualitative	Can handle undetected and false positions	Out of field balls, small database, size and shape changes

Table 8 continued

References	Sports	Ball detection	Ball trajectory generation	Dataset	Results	Strength	Shortcomings
Pallavi et al. (2008)	Soccer	Shape and size features, CHT	Directed weighted graph, dynamic programming	Real Madrid versus Manchester United at UEFA Champions League 2003, Chelsea versus Liverpool at UEFA Champions League 2007	Ball Detection—Medium shots : A—83.10%, Long shots : R—98.14, P—93.84.	Removed non-ball using static and dynamic features, pre-processing is not needed for filtering non-balls, efficient for poor playfield conditions	CHT detects only complete circles, thus proposed algorithm may miss non-ball looking ball candidate, ball may not look like ball in long shots

A accuracy, R recall, P precision

in tracking accuracy and robustness to occlusions. However, minor cases of misdetection were reported. Table 9 presents a summary of all Energy optimization based ball tracking methods published so far in different sports, validating databases, obtained results, strengths and shortcomings.

5 Discussion and our perspective

In this section, we present discussions on major ball detection and tracking techniques introduced so far followed by their classification and an improved model of a tracking system.

In most of the sports, a ball is a small object, highly agile, frequently occluded, changing contact and with changing contrast under varying illumination conditions. It is very important to detect the ball robustly, before or without tracking. Thus, all the approaches have used large number of features for its detection though it adds to computational complexity. Some of the preferred features of the ball consist of shape, size, color, position, motion, etc. A few curve detection methods like circle Hough transform, ellipse fitting and SVM multiple feature based classifiers were also used, resulting in a unified approach for detection unlike feature based detection.

Initially, most of the KF based approaches used a template to perform ball tracking. However, this technique proved inefficient in conditions of non-linear ball movement, illumination changes, occlusion, cluttered background and high speed of the ball. A large number of KF variants were preferred for ball tracking. They proved successful for dealing with size and time complexity but on a few very small databases with low-resolution video frames. In general, KF based method failed to incorporate non-linearity in their tracking techniques and needed manual input of initial ball location. Thus, a large number of researches used PF based methods for ball tracking. Increase in the number of assigned particles slowed down the tracking process. In contrast, reducing the number of particles increased the false positives and resulted in tracking failure. A balance between large number of particles and small number of false positives and negatives results in better tracking experience. This method was successful in tracking non-linear motion of the ball, redetection and partial occlusion; but could only marginally solve the problem of heavy occlusion, illumination changes, fast motion, motion blur, background clutter and size and scale changes. Both the probabilistic approaches need removal of salient objects on the ground like lines, circles, net, etc. for better ball detection.

Trajectory based approaches included mathematical relations of motion models and trajectories for tracking a ball. Some initial researchers used triangular geometric relations for detection of a ball in flight and eventually its trajectory. Trajectory based approaches, however, were mostly used for obtaining 2D and 3D trajectories rather than for complete tracking. A few of these methods were also used to convert the 2D trajectories into 3D. This conversion, however, proved costly in terms of computation and complexity. They even proved successful and yielded good results but the computation time for most of the works were not reported; we expect it to be higher than only feature based detection. This introduced ambiguity regarding the complexity of the system due to lack of complete information. The method of ball tracking based on 2D trajectory information required comparatively lesser computation time because there was no need to remove salient ground objects like borderlines, center circle, etc. unlike in the above filter based approaches and 3D techniques. Better visualization plots and prediction algorithms helped to obtain a more accurate trajectory in 2D and 3D space. However, most of the techniques did not consider ‘ball bounce’. Multiple

Table 9 Summary of Optimizing an Objective function based approach for Ball Tracking

References	Sports	Tracking methodology	Dataset	Result	Strength	Shortcomings
Wang et al. (2014a)	Soccer, Basketball	Focus-on-Scene tracker based on CRF loss function	FIBAW-1, FIBAW-2, APIDIS, ISSIA	Processing time 8–10 fps	Real-time	Few misdetections, physics based model not introduced
Wei et al. (2014)	Soccer, Tennis	a-HCRF	Soccer—9 complete matches of top-tier soccer league. Tennis—Australian Open Men's 2012 match of Novak Djokovic	NA	Future Prediction possible	Physical model not used
Wang et al. (2014b)	Soccer	MIP	FIBA	NA	Generic, tracks all objects simultaneously	Does not preserve identities throughout the sequence
Maksai et al. (2015)	Soccer, Volleyball, Basketball	MIP with physics based model	Basket-1, Basket-2, Basket-APIDIS, Volley-1, Volley-2, Soccer-ISSIA	NA	Generic, less unpredictability for shorter shots	Processing time increases for longer video sequences

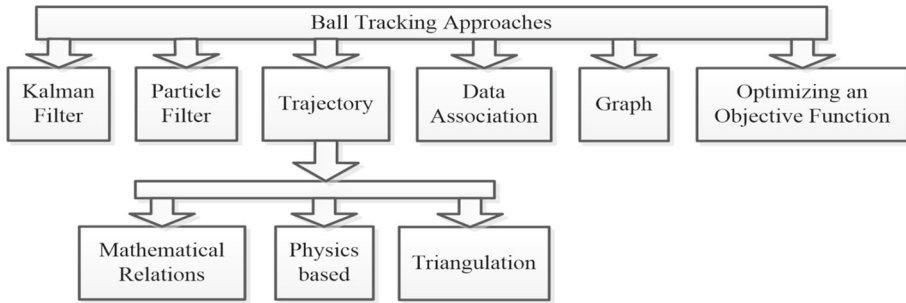


Fig. 9 Classification of ball tracking approaches

filters used in physics based approaches helped in reducing the number of non-ball candidates. None of the methods could satisfactorily tackle problems of changes in size, scale, color as well as cluttered background, air friction and high speed of the ball. Still, trajectory based approach could fairly deal with the basic problem of occlusion largely than KF and PF based approaches.

Initial methods treated DA as a motion-fitting problem and used RANSAC to solve it. However, it becomes computationally expensive in presence of background clutter and may result in failure of tracking. Multi-layered DA approaches are robust for abrupt motion change and incidental absence of the target object in frames. Layered DA methods made the tracking algorithm applicable for tracking multiple objects as well. They solved the problem of occlusion only in Tennis sports. DA based approach is prone to misdetections, cluttered background and fail to track ball temporarily going out of field. Similar to trajectory-based approach, DA based approach is an offline tracking method.

Like DA, Graph based approaches were also one of the choices of the researchers for handling short-term occlusion and redetection. However, most of the graph based approaches failed when either the ball was occluded or it got merged with players or line segments on the play field or moved out of field. Viterbi graph based algorithm was used for finding the optimal ball path. It loses track of target (ball) a few times because of lack of a well-defined motion model. This approach cannot handle high speed of the ball as well as changes in size, shape, color, illumination and appearance.

Traditional approach of ball detection and then tracking is a discrete approach with very less or no dependency between different stages of the framework. The latest approach of optimization of an objective energy function considers player-ball interactions for effective ball tracking. It converts the interactions between all the entities on court, into an objective function under physics based constraints. Objective functions modeled using conditional random fields (CRF), its variants and mixed integer program (MIP) were optimized. They produced excellent results in presence of occlusion in sports like, Soccer and Basketball. MIP makes the system real-time, accurate and robust to even long-term occlusion. This approach cannot handle high speed of the ball, false positives and abrupt ball motion. Unlike approaches discussed above, this generic approach can be used for similar other team sports as well.

We present classification overview of different ball tracking techniques in Fig. 9.

We further propose a modified block diagram for a ball tracking system with a feedback loop for learning current appearance of the ball as shown in Fig. 10.

Modern tracking systems continuously keep on updating the target template or appearance and even background parameters. This helps them learning in presence of various already

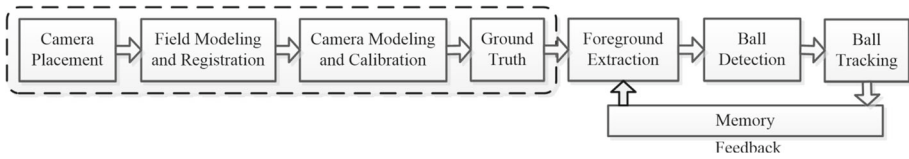


Fig. 10 Proposed model of an enhanced ball tracking system

said challenges and offer robust tracking performance. Available computationally powerful platforms, efficient learning and robust tracking algorithms are able to offer close to human and real-time performance to these systems. However, the capability of learning the target and background features is an important recent enhancement over the conventional tracking system architecture presented earlier in Fig. 1.

6 Conclusion and future scope

Availability of suitable cameras, their appropriate and correct installation, calibration, play-field model and game rule models are assumed available in all ball-tracking problems. The ball tracking problem is attacked in two steps; the first, detection of a ball and the second tracking the ball. The ball-tracking problem in sports has undergone many modifications due to the increased expectations; and a number of solutions have been offered. Features, shape, size, template and trajectory based approaches have been successfully proposed for ball detection. Constructing a deformable or dynamically changeable ball model in terms of different features under challenging conditions is going to be at the root of correct ball detection and subsequent tracking.

Ball Tracking techniques have evolved a lot over the years. It began with estimating 3D position of a ball in long trajectory using mathematical relations. However, it was a very crude method that needed reliable reference objects those are rarely available. The methods in those days considered ball tracking as an independent endeavor. The use of KF and then PF along with TM methods made the ball tracking process little bit robust to size and shape changes and short-term occlusions. However, it was still difficult to establish a track when the ball reappeared after long-term occlusion. Occlusion problem was mostly solved (but not completely) by Trajectory based methods. These methods were robust in obtaining missing and merged balls. However, most of them could not completely solve the problem of size and shape changes. Recently, DA methods have been introduced; but they were more confined to problem of ball tracking in Tennis that is considered solved. It could not handle the difficulties of other team sports like Soccer, Volleyball, Basketball, Cricket, etc. Graph based approaches used dynamic programming for obtaining the track of the ball during the game. However, this approach could not do wonders, as there was nothing new in the framework. Recently introduced approach of optimizing an objective function yielded impressive results as it took player and ball interactions into consideration and employed physics based constraints thereby predicting the ball position and successfully tracking it. This is a generic framework and can be applied for other team sports where major problems are occlusion and rapid player-ball movements. However, it is not fast enough for commercially viable use for a sports broadcaster. The future lies in the right employment of deep learning techniques where artificial intelligence predicts the future ball movements using knowledge from previous frames and matches. More accurate estimation of trajectory, may be possible by modelling effects like air friction, ball spin, complex ball movements like swerve, etc.

The worldwide use of Hawkeye Technologies in all sports has proven that even simple theories as triangulation techniques combined with capabilities of high-speed high resolution multiple cameras can provide solutions to most of the problems including heavy occlusion in ball tracking. There is also a scope to develop newer theories and methods to handle problems of heavy or long-term occlusion, abrupt motion changes and tracking and trajectory computations in 3D space. New and precise databases have to be introduced in order to evaluate and benchmark the tracking algorithms. The available databases are very few and designed only for sports like football (ISSIA) and basketball (APIDIS, FIBAW). Others sports, as cricket also should have its relevant databases and ground truths under different possible challenging conditions. Multiple Balls tracking with similar features like color and shape in games like snooker and billiards is a much more complex task.

The ball tracking algorithms introduced for different sports cannot be compared on a common platform. Requirements, expectations, physical infrastructure and situations of interest in every sport are different. The quantitative performance benchmarking is also difficult due to unavailability of suitable databases and ground truths for most of the sports. The Broadcast videos used for experimental purpose may have large variations with respect to the coding parameters and type, resolution, frame rate, etc. Even the videos taken by fixed or moving cameras could not achieve consensus for experimentation because of different types of video capturing devices used, their parameter variations, installation, etc. All these points lead to difficulties in building up a universal concrete infrastructure for ball tracking systems; however, it is very well possible for a specific sport. This exhaustive study makes it possible to identify advantages, features, limitations, infrastructure requirements and quantitative performance benchmarking for the published algorithms and subsequently for the infrastructure.

References

- Almajai I, Yan F, de Campos T, Khan A, Christmas W, Windridge D, Kittler J (2012) Anomaly detection and knowledge transfer in automatic sports video annotation. In: Detection and identification of rare audiovisual cues. Springer, pp 109–117
- Ancona N, Cicirelli G, Stella E, Distante A (2003) Ball detection in static images with support vector machines for classification. *Image Vis Comput* 21:675–692
- Ariki Y, Takiguchi T, Yano K (2008) Digital camera work for soccer video production with event recognition and accurate ball tracking by switching search method. In: 2008 IEEE international conference on multimedia and expo. IEEE, pp 889–892
- Arulampalam MS, Maskell S, Gordon N, Clapp T (2002) A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans Signal Process* 50:174–188
- Bar-Shalom Y (1987) Tracking and data association. Academic Press, London
- Barnard M, Odobez J-M (2004) Robust playfield segmentation using MAP adaptation. In: Proceedings of the 17th international conference on pattern recognition. ICPR 2004. IEEE, pp 610–613
- Batz G, Lee K-K, Wollherr D, Buss M (2009) Robot basketball: a comparison of ball dribbling with visual and force/torque feedback. In: IEEE international conference on robotics and automation. ICRA'09. IEEE, pp 514–519
- Beetz M, von Hoyningen-Huene N, Kirchlechner B, Gedikli S, Siles F, Durus M, Lames M (2009) Aspogamo: automated sports game analysis models. *Int J Comput Sci Sport* 8:1–21
- Birbach O, Frese UA (2009) Multiple hypothesis approach for a ball tracking system. In: International conference on computer vision systems. Springer, pp 435–444
- Birbach O, Frese U, Bäuml B (2011) Realtime perception for catching a flying ball with a mobile humanoid. In: 2011 IEEE international conference on robotics and automation (ICRA). IEEE, pp 5955–5962
- Bloom T, Bradley AP (2003) Player tracking and stroke recognition in tennis video. In: APRS workshop on digital image computing (WDIC'03). The University of Queensland, pp 93–97
- Chakraborty B, Meher S (2011) 2D trajectory-based position estimation and tracking of the ball in a basketball video. In: Second international conference on trends in optics and photonics, Kolkata, India, December 7–9, 2011, pp 537–545

- Chakraborty B, Meher S (2012a) Real-time position estimation and tracking of a basketball. In: 2012 IEEE international conference on signal processing, computing and control (ISPCC). IEEE, pp 1–6
- Chakraborty B, Meher S A (2012b) Trajectory-based ball detection and tracking system with applications to shot-type identification in volleyball videos. In: 2012 international conference on signal processing and communications (SPCOM). IEEE, pp 1–5
- Chakraborty B, Meher S (2013a) A real-time trajectory-based ball detection-and-tracking framework for basketball video. *J Opt* 42:156–170
- Chakraborty B, Meher S A (2013b) Trajectory-based ball detection and tracking system with applications to shooting angle and velocity estimation in basketball videos. In: 2013 Annual IEEE India conference (INDICON). IEEE, pp 1–6
- Chang M-H, Tien M-C, Wu J-L (2009) WOW: wild-open warning for broadcast basketball video based on player trajectory. In: Proceedings of the 17th ACM international conference on multimedia. ACM, pp 821–824
- Chen B, Wang Z (2007) A statistical method for analysis of technical data of a badminton match based on 2-D seriate images. *Tsinghua Sci Technol* 12:594–601
- Chen H-T, Chen H-S, Hsiao M-H, Tsai W-J, Lee S-Y (2008) A trajectory-based ball tracking framework with visual enrichment for broadcast baseball videos. *J Inf Sci Eng* 24:143–157
- Chen H-T, Chen H-S, Lee S-Y (2007) Physics-based ball tracking in volleyball videos with its applications to set type recognition and action detection. In: 2007 IEEE international conference on acoustics, speech and signal processing-ICASSP'07. IEEE, pp 1-1097–1-1100
- Chen H-T, Tien M-C, Chen Y-W, Tsai W-J, Lee S-Y (2009) Physics-based ball tracking and 3D trajectory reconstruction with applications to shooting location estimation in basketball video. *J Vis Commun Image Represent* 20:204–216
- Chen H-T, Tsai W-J, Lee S-Y, Yu J-Y (2012) Ball tracking and 3D trajectory approximation with applications to tactics analysis from single-camera volleyball sequences. *Multimed Tools Appl* 60:641–667
- Chen W, Zhang Y-J (2006) Tracking ball and players with applications to highlight ranking of broadcasting table tennis video. In: IMACS multiconference on computational engineering in systems applications. IEEE, pp 1896–1903
- Cheng X, Honda M, Ikoma N, Ikenaga T (2016) Anti-occlusion observation model and automatic recovery for multi-view ball tracking in sports analysis. In: 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 1501–1505
- Cheng X, Zhuang X, Wang Y, Honda M, Ikenaga T (2015) Particle filter with ball size adaptive tracking window and ball feature likelihood model for ball's 3D position tracking in volleyball analysis. In: Pacific Rim conference on multimedia. Springer, pp 203–211
- Choi K, Seo Y (2004) Probabilistic tracking of the soccer ball. In: International workshop on statistical methods in video processing. Springer, pp 50–60
- Choi K, Seo Y (2005) Tracking soccer ball in TV broadcast video. In: International conference on image analysis and processing. Springer, pp 661–668
- Choi K, Seo Y (2011) Automatic initialization for 3D soccer player tracking. *Pattern Recogn Lett* 32:1274–1282
- Christmas W, Kolonias I, Kittler J, Yan F (2007) Improving the accuracy of automatic tennis video annotation by high level grammar. In: 14th international conference on image analysis and processing workshops. ICIAPW 2007. IEEE, pp 154–159
- Cigliano P, Lippiello V, Ruggiero F, Siciliano B (2015) Robotic ball catching with an eye-in-hand single-camera system. *IEEE Trans Control Syst Technol* 23:1657–1671
- Conaire CÓ, Kelly P, Connaghan D, O'Connor NE (2009) Tennissense: a platform for extracting semantic information from multi-camera tennis data. In: 2009 16th international conference on digital signal processing. IEEE, pp 1–6
- D'Orazio T, Guaragnella C, Leo M, Distante A (2004) A new algorithm for ball recognition using circle Hough transform and neural classifier. *Pattern Recogn* 37:393–408
- D'Orazio T, Leo M, Spagnolo P, Mazzeo PL, Mosca N, Nitti M, Distante A (2009) An investigation into the feasibility of real-time soccer offside detection from a multiple camera system. *IEEE Trans Circuits Syst Video Technol* 19:1804–1818
- D'Orazio T, Leo M, Spagnolo P, Nitti M, Mosca N, Distante A (2009) A visual system for real time detection of goal events during soccer matches. *Comput Vis Image Underst* 113:622–632
- Dearden A, Demiris Y, Grau O (2006) Tracking football player movement from a single moving camera using particle filters. In: Proceedings of the 3rd European conference on visual media production (CVMP), London, pp 29–37

- Desai UB, Merchant SN, Zaveri M, Ajishna G, Purohit M, Phanish H (2005) Small object detection and tracking: algorithm, analysis and application. In: International conference on pattern recognition and machine intelligence. Springer, pp 108–117
- Ekinci B, Gokmen MA (2008) Ball tracking system for offline tennis videos. In: International conference on visualization, imaging and simulation
- El Abed A, Dubuisson S, Béréziat D (2006) Comparison of statistical and shape-based approaches for non-rigid motion tracking with missing data using a particle filter. In: International conference on advanced concepts for intelligent vision systems. Springer, pp 185–196
- Farin D, de Han J (2005) With PH fast camera calibration for the analysis of sport sequences. In: 2005 IEEE international conference on multimedia and expo. IEEE
- Farin D, de Krabbe S, Withb PH, Effelsberga W (2004) Robust camera calibration for sport videos using court models. In: SPIE proceedings series, 2004. Society of Photo-Optical Instrumentation Engineers, pp 80–91
- Figuerola P, Leite N, Barros RM, Cohen I, Medioni G (2004) Tracking soccer players using the graph representation. In: Proceedings of the 17th international conference on pattern recognition. ICPR 2004. IEEE, pp 787–790
- Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM* 24:381–395
- Glover J, Kaelbling LP (2014) Tracking the spin on a ping pong ball with the quaternion Bingham filter. In: 2014 IEEE international conference on robotics and automation (ICRA). IEEE, pp 4133–4140
- Han M, Hua W, Chen T, Gong Y (2003) Feature design in soccer video indexing. In: Proceedings of the 2003 joint conference of the fourth international conference on information, communications and signal processing, 2003 and Fourth Pacific Rim conference on multimedia. IEEE, pp 950–954
- Hartley R, Zisserman A (2003) Multiple view geometry in computer vision. Cambridge University Press, Cambridge
- Hossein-Khani J, Soltanian-Zadeh H, Kamarei M, Staadt O (2011) Ball detection with the aim of corner event detection in soccer video. In: 2011 Ninth IEEE international symposium on parallel and distributed processing with applications workshops (ISPAW). IEEE, pp 147–152
- Hoyningen-Huene Nv, Beetz M (2009) Rao-blackwellized resampling particle filter for real-time player tracking in sports VISAPP, INSTICC, pp 464–471
- Hu M-C, Chang M-H, Wu J-L, Chi L (2011) Robust camera calibration and player tracking in broadcast basketball video. *IEEE Trans Multimed* 13:266–279
- Huang Q, Cox S, Yan F, deCampos T, Windridge D, Kittler J, Christmas W (2011) Improved detection of ball hit events in a tennis game using multimodal information. In: 11th International conference on auditory-visual speech processing (AVSP)
- Huang Q, Cox S, Zhou X, Xie L (2012) Detection of ball hits in a tennis game using audio and visual information. In: 2012 Asia-Pacific signal and information processing association annual summit and conference (APSIPA ASC). IEEE, pp 1–10
- Huang Y, Llach J, Zhang C A (2008) method of small object detection and tracking based on particle filters. In: 19th international conference on pattern recognition. ICPR 2008. IEEE, pp 1–4
- Ishii N, Kitahara I, Kameda Y, Ohta Y (2007) 3D tracking of a soccer ball using two synchronized cameras. In: Pacific-Rim conference on multimedia. Springer, pp 196–205
- Iwase S, Saito H (2003) Tracking soccer players based on homography among multiple views. In: Visual communications and image processing. International Society for Optics and Photonics, pp 283–292
- Janssen R, Verrijt M, de Best J, van de Molengraft R (2012) Ball localization and tracking in a highly dynamic table soccer environment. *Mechatronics* 22:503–514
- Kasiri-Bidhendi S, Safabakhsh R (2009) Effective tracking of the players and ball in indoor soccer games in the presence of occlusion. In: 14th International CSI computer conference. CSICC 2009. IEEE, pp 524–529
- Kasuya N, Kitahara I, Kameda Y, Ohta Y (2010) Real-time soccer player tracking method by utilizing shadow regions. In: Proceedings of the 18th ACM international conference on Multimedia. ACM, pp 1319–1322
- Kelly P et al (2010) A low-cost performance analysis and coaching system for tennis. In: Proceedings of ACM multimedia
- Kim J-Y, Kim T-Y (2009) Soccer ball tracking using dynamic kalman filter with velocity control. In: Sixth international conference on computer graphics, imaging and visualization. CGIV'09. IEEE, pp 367–374
- Kim T, Seo Y, Hong K-S (1998) Physics-based 3D position analysis of a soccer ball from monocular image sequences. In: Sixth international conference on computer vision. IEEE, pp 721–726
- Kittler J, Christmas WJ, Kostin A, Yan F, Kolonias I, Windridge DA (2005) memory architecture and contextual reasoning framework for cognitive vision. In: Scandinavian conference on image analysis. Springer, pp 343–358

- Kokaram A, Pitie F, Dahyot R, Rea N, Yeterian S (2005) Content controlled image representation for sports streaming. In: Proceedings of content-based multimedia indexing (CBMI'05)
- Kumar A, Chavan PS, Sharatchandra V, David S, Kelly P, O'Connor NE (2011) 3D estimation and visualization of motion in a multicamera network for sports. In: 2011 Irish machine vision and image processing conference (IMVIP). IEEE, pp 15–19
- Leo M, Mosca N, Spagnolo P, Mazzeo PL, D'Orazio T, Distanto A (2008) Real-time multiview analysis of soccer matches for understanding interactions between ball and players. In: Proceedings of the 2008 international conference on content-based image and video retrieval. ACM, pp 525–534
- Lepetit V, Shahroki A, Fua P (2003) Robust data association for online application. In: Proceedings. IEEE Computer Society conference on computer vision and pattern recognition. IEEE, pp I–I
- Li Y, Dore A, Orwell J (2005) Evaluating the performance of systems for tracking football players and ball. In: IEEE conference on advanced video and signal based surveillance. IEEE, pp 632–637
- Liang D, Huang Q, Liu Y, Zhu G, Gao W (2007) Video2Cartoon: a system for converting broadcast soccer video into 3D cartoon animation. IEEE Trans Consum Electron 53:1138–1146
- Liang D, Liu Y, Huang Q, Gao WA (2005) Scheme for ball detection and tracking in broadcast soccer video. In: Pacific-Rim conference on multimedia. Springer, pp 864–875
- Liu J, Fang Z, Zhang K, Tan M (2014) Improved high-speed vision system for table tennis robot. In: 2014 IEEE international conference on mechatronics and automation. IEEE, pp 652–657
- Liu Y, Liang D, Huang Q, Gao W (2006) Extracting 3D information from broadcast soccer video. Image Vis Comput 24:1146–1162
- Ltd. H-EI (2015) Ball tracking. <https://www.hawkeyeinnovations.com/products/ball-tracking>. Accessed 10 July 2017
- Maksai A, Wang X, Fua P (2015) What players do with the ball: a physically constrained interaction modeling. arXiv preprint [arXiv:1511.06181](https://arxiv.org/abs/1511.06181)
- Mauthner T, Koch C, Tilp M, Bischof H (2007) Visual tracking of athletes in beach volleyball using a single camera. Int J Comput Sci Sport 6:21–34
- Misu T, Matsui A, Naemura M, Fujii M, Yagi N (2007) Distributed particle filtering for multiocular soccer-ball tracking. In: 2007 IEEE international conference on acoustics, speech and signal processing-ICASSP'07, 2007. IEEE, pp III-937–III-940
- Miura J, Shimawaki T, Sakiyama T, Shirai Y (2009) Ball route estimation under heavy occlusion in broadcast soccer video. Comput Vis Image Underst 113:653–662
- Miyamori H, Iisaku S-I (2000) Video annotation for content-based retrieval using human behavior analysis and domain knowledge. In: Fourth IEEE international conference on automatic face and gesture recognition. Proceedings. IEEE, pp 320–325
- Naidoo WC, Tapamo JR (2006) Soccer video analysis by ball, player and referee tracking. In: Proceedings of the 2006 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries, 2006. South African Institute for Computer Scientists and Information Technologists, pp 51–60
- Ohno Y, Miura J, Shirai Y (2000) Tracking players and estimation of the 3D position of a ball in soccer games. In: 15th international conference on pattern recognition. Proceedings. IEEE, pp 145–148
- Ohno Y, Miura J, Shirai Y (1999) Tracking players and a ball in soccer games. In: IEEE/SICE/RSJ international conference on multisensor fusion and integration for intelligent systems. MFI'99. Proceedings. IEEE, pp 147–152
- Pallavi V, Mukherjee J, Majumdar AK, Sural S (2008) Ball detection from broadcast soccer videos using static and dynamic features. J Vis Commun Image Represent 19:426–436
- Pei C, Yang S, Gao L, Ma W (2009) A real time ball detection framework for soccer video. In: 2009 16th international conference on systems, signals and image processing. IEEE, pp 1–4
- Perše M, Kristan M, Perš J, Vuckovic G, Kovacic S (2005) Physics-based modelling of human motion using Kalman filter and collision avoidance algorithm. In: International symposium on image and signal processing and analysis, ISPA05, Zagreb, Croatia. Citeseer, pp 328–333
- Pingali G, Opalach A, Jean Y (2000) Ball tracking and virtual replays for innovative tennis broadcasts. In: 15th International conference on pattern recognition. Proceedings. IEEE, pp 152–156
- Pingali GS, Opalach A, Jean YD, Carlbom IB (2002) Instantly indexed multimedia databases of real world events. IEEE Trans Multimed 4:269–282
- Poliakov A, Marraud D, Reithler L, Chatain C (2010) Physics based 3D ball tracking for tennis videos. In: 2010 international workshop on content-based multimedia indexing (CBMI). IEEE, pp 1–6
- Rea N, Dahyot R, Kokaram A (2004) Semantic event detection in sports through motion understanding. In: International conference on image and video retrieval. Springer, pp 88–97
- Reid ID, North A (1998) 3D Trajectories from a single viewpoint using shadows. In: BMVC, pp 51–52
- Ren J, Orwell J, Jones G (2006) Generating ball trajectory in soccer video sequences

- Ren J, Orwell J, Jones G, Xu M (2004a) Real-time 3D soccer ball tracking from multiple cameras. In: Proceedings of the British Machine vision conference (BMVC'04), pp 829–838
- Ren J, Orwell J, Jones GA, Xu M (2004b) A general framework for 3D soccer ball estimation and tracking. In: 2004 international conference on image processing. ICIP'04. IEEE, pp 1935–1938
- Ren J, Orwell J, Jones GA, Xu M (2008) Real-time modeling of 3-d soccer ball trajectories from multiple fixed cameras. *IEEE Trans Circuits Syst Video Technol* 18:350–362
- Ren J, Orwell J, Jones GA, Xu M (2009) Tracking the soccer ball using multiple fixed cameras. *Comput Vis Image Underst* 113:633–642
- Santiago CB, Sousa A, Reis LP, Estriga ML (2011) Real time colour based player tracking in indoor sports. In: Computational vision and medical image processing. Springer, pp 17–35
- Seo Y, Choi S, Kim H, Hong K-S (1997) Where are the ball and players? Soccer game analysis with color-based tracking and image mosaick. In: International conference on image analysis and processing. Springer, pp 196–203
- Shah H, Chokalingam P, Paluri B, Pradeep N, Raman B (2007) Automated stroke classification in tennis. In: International conference image analysis and recognition. Springer, pp 1128–1137
- Shimawaki T, Miura J, Sakiyama T, Shirai Y (2006a) Ball route estimation in broadcast soccer video. In: Proceedings of ECCV-2006 workshop on computer vision based analysis in sport environments. Citeseer, pp 26–37
- Shimawaki T, Sakiyama T, Miura J, Shirai Y (2006b) Estimation of ball route under overlapping with players and lines in soccer video image sequence. In: 18th international conference on pattern recognition (ICPR'06). IEEE, pp 359–362
- Shum H, Komura TA (2004) Spatiotemporal approach to extract the 3D trajectory of the baseball from a single view video sequence. In: 2004 IEEE international conference on multimedia and expo. ICME'04. IEEE, pp 1583–1586
- Shum H, Komura T (2005) Tracking the translational and rotational movement of the ball using high-speed camera movies. In: IEEE international conference on image processing. IEEE, pp III-1084–III-1087
- Teachabarikiti K, Chalidabhongse TH, Thammano A (2010) Players tracking and ball detection for an automatic tennis video annotation. In: 2010 11th international conference on control automation robotics and vision (ICARCV). IEEE, pp 2461–2494
- Tong X-F, Lu H-Q, Liu Q-S (2004) An effective and fast soccer ball detection and tracking method. In: Proceedings of the 17th international conference on pattern recognition. ICPR 2004. IEEE, pp 795–798
- Tong X, Liu J, Wang T, Zhang Y (2011) Automatic player labeling, tracking and field registration and trajectory mapping in broadcast soccer video. *ACM Trans Intell Syst Technol* 2:15
- Uchiyama H, Saito H (2007) AR display of visual aids for supporting pool games by online markerless tracking. In: 17th international conference on artificial reality and teleexistence. IEEE, pp 172–179
- Wang X, Ablavsky V, Shritit HB, Fua P (2014a) Take your eyes off the ball: improving ball-tracking by focusing on team play. *Comput Vis Image Underst* 119:102–115
- Wang X, Türetken E, Fleuret F, Fua P (2014b) Tracking interacting objects optimally using integer programming. In: European conference on computer vision. Springer, pp 17–32
- Wang X, Türetken E, Fleuret F, Fua P (2016) Tracking interacting objects using intertwined flows. *IEEE Trans Pattern Anal Mach Intell* 38:2312–2326
- Wei X, Lucey P, Vidas S, Morgan S, Sridharan S (2014) Forecasting events using an augmented hidden conditional random field. In: Asian conference on computer vision. Springer, pp 569–582
- Wong K, Dooley L (2010) High-motion table tennis ball tracking for umpiring applications. In: IEEE 10th international conference on signal processing proceedings. IEEE, pp 2460–2463
- Xing J, Ai H, Liu L, Lao S (2011) Multiple player tracking in sports video: a dual-mode two-way bayesian inference approach with progressive observation modeling. *IEEE Trans Image Process* 20:1652–1667
- Yamada A, Shirai Y, Miura J (2002) Tracking players and a ball in video image sequence and estimating camera parameters for 3D interpretation of soccer games. In: 16th international conference on pattern recognition. Proceedings. IEEE, pp 303–306
- Yan F, Christmas W, Kittler J (2005) A tennis ball tracking algorithm for automatic annotation of tennis match. In: British machine vision conference, pp 619–628
- Yan F, Christmas W, Kittler J (2006a) A maximum a posteriori probability viterbi data association algorithm for ball tracking in sports video. In: 18th international conference on pattern recognition. ICPR 2006. IEEE, pp 279–282
- Yan F, Christmas W, Kittler J (2008) Layered data association using graph-theoretic formulation with application to tennis ball tracking in monocular sequences. *IEEE Trans Pattern Anal Mach Intell* 30:1814–1830
- Yan F, Christmas W, Kittler J (2014) Ball tracking for tennis video annotation. In: Computer vision in sports. Springer, pp 25–45

- Yan F, Kostin A, Christmas W, Kittler J (2006) A novel data association algorithm for object tracking in clutter with application to tennis video analysis. In: 2006 IEEE Computer Society conference on computer vision and pattern recognition (CVPR'06). IEEE, pp 634–641
- Yow D, Yeo B-L, Yeung M, Liu B (1995) Analysis and presentation of soccer highlights from digital video. In: Proceedings ACCV. Citeseer, pp 499–503
- Yu J, Tang Y, Wang Z, Shi L (2007a) Playfield and ball detection in soccer video. In: Advances in visual computing, pp 387–396
- Yu X, Hay TS, Yan X, Chng E (2005) A player-possession acquisition system for broadcast soccer video. In: 2005 IEEE international conference on multimedia and expo. IEEE, pp 522–525
- Yu X, Jiang N, Cheong L-F, Leong HW, Yan X (2009) Automatic camera calibration of broadcast tennis video with applications to 3D virtual content insertion and ball detection and tracking. *Comput Vis Image Underst* 113:643–652
- Yu X, Leong HW, Xu C, Tian Q (2004a) A robust and accumulator-free ellipse Hough transform. In: Proceedings of the 12th annual ACM international conference on multimedia. ACM, pp 256–259
- Yu X, Leong HW, Xu C, Tian Q (2006) Trajectory-based ball detection and tracking in broadcast soccer video. *IEEE Trans Multimed* 8:1164–1178
- Yu X, Sim C-H, Wang JR, Cheong LF (2004) A trajectory-based ball detection and tracking algorithm in broadcast tennis video. In: 2004 international conference on image processing. ICIP'04. IEEE, pp 1049–1052
- Yu X, Tian Q, Wan KW (2003a) A novel ball detection framework for real soccer video. In: 2003 international conference on multimedia and expo. ICME'03. Proceedings. IEEE, vol 262, pp II-265–II-268
- Yu X, Tu X, Ang EL (2007b) Trajectory-based ball detection and tracking in broadcast soccer video with the aid of camera motion recovery. In: 2007 IEEE international conference on multimedia and expo. IEEE, pp 1543–1546
- Yu X, Xu C, Leong HW, Tian Q, Tang Q, Wan KW (2003b) Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video. In: Proceedings of the eleventh ACM international conference on multimedia. ACM, pp 11–20
- Yu X, Xu C, Tian Q, Leong HW (2003c) A ball tracking framework for broadcast soccer video. In: 2003 international conference on multimedia and expo. ICME'03. Proceedings. IEEE, vol 272, pp II-273–II-276
- Yu X, Xu C, Tian Q, Yan X, Wan KW, Jiang Z (2003d) Estimation of the ball size in broadcast soccer video using salient objects. In: Proceedings of the 2003 joint conference of the fourth international conference on information, communications and signal processing, 2003 and Fourth Pacific Rim conference on multimedia. IEEE, pp 930–934
- Yu X, Yan X, Hay TS, Leong HW (2004c) 3D reconstruction and enrichment of broadcast soccer video. In: Proceedings of the 12th annual ACM international conference on multimedia. ACM, pp 260–263
- Zaveri MA, Merchant SN, Desai UB (2004) Small and fast moving object detection and tracking in sports video sequences. In: 2004 IEEE international conference on multimedia and expo. ICME'04. IEEE, pp 1539–1542
- Zhang Y-h, Wei W, Yu D, Zhong C-w (2011) A tracking and predicting scheme for ping pong robot. *J Zhejiang Univ Sci C* 12:110–115
- Zhang Z, Xu D, Tan M (2010) Visual measurement and prediction of ball trajectory for table tennis robot. *IEEE Trans Instrum Meas* 59:3195–3205
- Zhou X, Huang Q, Xie L, Cox S (2013) A two layered data association approach for ball tracking. In: 2013 IEEE international conference on acoustics, speech and signal processing. IEEE, pp 2317–2321
- Zhou X, Xie L, Huang Q, Cox SJ, Zhang Y (2015) Tennis ball tracking using a two-layered data association approach. *IEEE Trans Multimed* 17:145–156
- Zhu G, Xu C, Huang Q, Rui Y, Jiang S, Gao W, Yao H (2009) Event tactic analysis based on broadcast sports video. *IEEE Trans Multimed* 11:49–67
- Zhu G, Xu C, Zhang Y, Huang Q, Lu H (2008) Event tactic analysis based on player and ball trajectory in broadcast video. In: Proceedings of the 2008 international conference on content-based image and video retrieval. ACM, pp 515–524