

Real-time 3D Soccer Ball Tracking from Multiple Cameras

Jinchang Ren, James Orwell, Graeme Jones, Ming Xu
Digital Imaging Research Centre, Kingston University
Surrey, KT1 2EE, U. K.

{j.ren, j.orwell, g.jones, m.xu}@kingston.ac.uk

Abstract

A novel model for 3D soccer ball tracking is proposed, using multiple fixed cameras as input. The main problems are: successfully filtering false alarms, tracking through missing observations and estimating 3D positions from single or multiple camera inputs. The key innovation is to categorise all motion of the ball into four possible phases – *rolling*, *flying*, *in possession*, and *out of play*, and in different phases the ball trajectory is further modelled as linear or curve segments. Then, triangulations and phase-specific models are employed to estimate 3D ball positions. The system accuracy is evaluated by comparing the estimated ball phases and positions with manual ground-truth, over several minutes of real data sequences. Unlike existing systems using shadows and manual assistance, our approach demonstrates fully automatic ball tracking with the potential for accurate and robust results.

1 Introduction

As a combination of computer vision and multimedia technologies, many important applications have been developed in automatic sports video analysis, especially in football video analysis [1-2]. These applications provide additional information for better comprehension of football games, such as video content annotation and summarization, verification of referee decisions and further 2D/3D reconstruction and visualization [1-10].

Accurate localisation of the positions of players and the ball has several uses, from post-match analysis of strategy and fitness to additional low-bandwidth streaming of novel spectator content. Naturally, the ball is an important component of this content. Although colour and template matching methods have been successfully used in detection and tracking of players and referees [1, 4, 7], they are less effective for ball detection and tracking due to several reasons [13]: the ball is smaller, often confused with parts of players and line markings, suffers rapid and unpredictable 3D acceleration.

Thus, the problem under investigation is the automatic estimation of ball position from multiple fixed cameras. Complete solutions for this problem are not currently

This work forms part of the INMOVE project, supported by the European Commission IST 2001-37422.

available in the literature. In the football (soccer) domain, fully automatic methods for limited scene understanding have been proposed, e.g. recognition of replays from cinematic features extracted from broadcast TV data [1], *detection* of the ball in broadcast TV data [5], and limited ball *tracking* for selection of optimal viewpoint from fixed cameras [3]. Also, sophisticated motion models have been proposed in conjunction with semi manual methods for input of measurements: in [7], the 3D ball trajectory is modelled by considering air friction and gravity but depends on an unsolved initial velocity. *Bebie* and *Bieri* model 3D trajectory segments by Hermite spline curves [2], yet the ball positions in about every fifth frames need manual identification. Also, in [6] and [11], reference players and shadows are used as additional cues to estimate of 3D ball positions. These additional cues are not required for the work presented below.

In this paper, the models and estimators we used are not as sophisticated as some in the literature. Rather, the novel contributions are: the definition, recognition, and modelling of different *phases* of ball motion – namely *rolling*, *flying*, *in-possession*, and *out-of-play*, and also 3D ball positioning using triangulations and phase-specific models. The phase transitions in ball motion are useful for several reasons. Firstly, it provides a helpful means of constraining the model to limit the uncertainty. Secondly, it allows further cues to be adopted for recognising the actual motion phases. Finally, it in itself provides a useful form of output.

The approach is as follows: Image data from multiple cameras is processed separately to identify the moving objects, and each object is attributed a measure of the likelihood of being a ball. From all the cameras, the data about these objects is used to estimate the most probable *phase* of ball motion. In each of the phase transition, the ball trajectory is modelled as linear or curve segment. Triangulations and phase-specific models are then employed to estimate 3D ball positions and successive phase transitions.

2 Detecting Ball-like Features

2.1 Detecting and Tracking Moving Objects

Image differencing and Kalman-based tracking are applied to generate observations of objects moving in the field of view (FOV) of a single camera. Here, an adaptive background model, consisting of a per-pixel mixture-of-Gaussians model [12], is used in our image differencing operation. Its output is ‘regions’ of connected components, and each region i is represented by image plane observation $\mathbf{o}_i = [r_0 \ c_0 \ r_1 \ c_1 \ r_2 \ c_2]^T$, where (r_0, c_0) is the centroid, (r_1, c_1) and (r_2, c_2) are the top-left and bottom-right corners of the bounding box, respectively.

A simple image plane tracker [9] is used to filter noisy measurements and split merged objects. The measurements are represented in common ground plane co-ordinates, using the Tsai’s method of camera calibration. Until Section 3, all objects are assumed to lie on the ground plane (usually true for players, but the ball could be anywhere on the line between that point and the camera position). A ground-plane measurement $\mathbf{m}_i = [x_w \ y_w \ z_w \ w \ h \ a]^T$ is defined, where (x_w, y_w, z_w) is the 3D object position (with initial estimate $z_w = 0$), and w, h and a are *width*, *height* and *area* of the object (also expressed in ground plane co-ordinates, assuming it is touching the ground

plane). Figure 1 plots trajectories in (t, c_0) format from frame 950 to 2100 in camera sequence 1, in which frame number t and image plane co-ordinate c_0 is used.

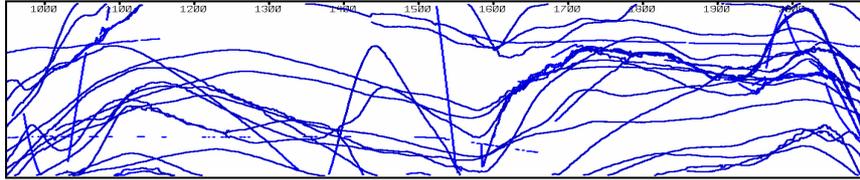


Figure 1: Thirty seconds of single camera tracking data (Time t moves from left to right, and the x -coordinate of the objects c_0 is plotted up the y -axis of the Figure)

2.2 Identifying Ball-like Features

The final step in the single-view process is to attribute each of the ground plane measurements \mathbf{m}_i with likelihood $l(\mathbf{m}_i)$ that it represents the ball. The other objects represented by the \mathbf{m}_i are: the players and referees, moving crowd, and field markings erroneously classed as foreground (due to camera shake or luminosity change).

The naive method of looking for small white objects is to identify an object as a likely ball candidate only if its *width*, *height* and/or *area* as well as percentage of pixels in white lie within several given thresholds. This is compromised by several problems, which we identify and correct for. The first problem is that segmented field lines (such as those shown in Fig 2) are also small and white. The second problem is that fragments of players, especially socks, are erroneously identified as separate small white objects. Third, at this stage its 3D position is still unavailable: its size is calculated by assuming it is touching the ground plane. Thus, the airborne ball may appear to be a large object on the ground plane. Finally, the image of fast-moving ball is affected by motion blur, rendering it larger and less white than a stationary (or slower moving) one.

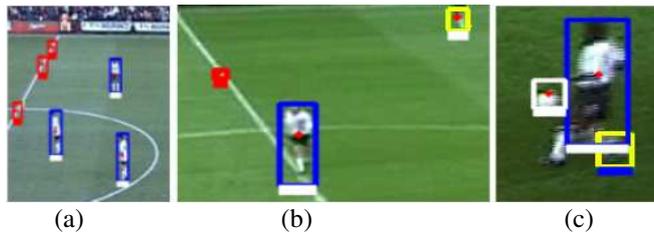


Figure 2: Enlarged images of detected moving objects in different colour boxes with the ball (in white), players (in blue), and false alarms (in red or yellow).

In our ball filtering process, *velocity* and *longevity* features are introduced to discriminate the ball from these other objects along with *size* and *colour* information. These features are employed in the simple heuristic method described below. A more sophisticated method (such as linear discriminant analysis or state vector machines) could be employed to greater effect.

The majority of white field markings of small size and low speed can be discriminated from the ball by including thresholds for both the size and absolute velocity of the detected object. Parts of players are eliminated by discarding all short-lived objects and merged them with players accompanied nearby. Cases when the ball is calculated to be larger than it really is (either through motion blur or incorrectly

assuming the object lies on the ground plane) are treated by making the thresholds (for size and percentage of pixels in ball colour) as functions of the estimated 2D object speed, thereby admitting fast-moving objects.

Indeed, the object speed has been found to be the most useful discriminant of all features. After thresholding using *size*, *velocity* and *colour* features, the filtered ball candidate are specified with ball likelihood of 0.5. To obtain our floating point measure of likelihood l_i ranging from 0 to 1, we use two parameters k_1 and k_2 to incorporate the velocity $\|\mathbf{v}\|$ and longevity n of the track into the likelihood measure as follows:

$$l_i = \frac{1}{2} + k_1 \cdot \frac{\|\mathbf{v}_i\| - \|\bar{\mathbf{v}}\|}{\|\bar{\mathbf{v}}\|} + k_2 \cdot n_i \quad (1)$$

where $\bar{\mathbf{v}}$ is the average velocity vector of all detected objects. Figure 3 plots ball filtering results using *size*, *velocity*, *colour* and *longevity* features.

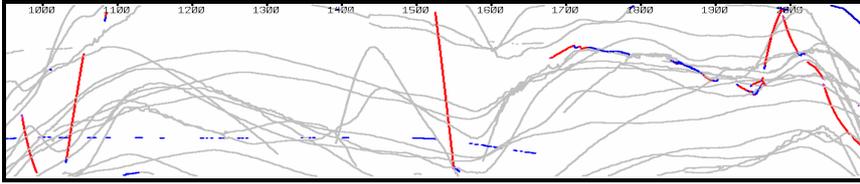


Figure 3: Ball filtering results of Fig. 1 using size, velocity, colour and longevity features to indicate the ball (labelled in red)

After ball filtering, the ball candidates are extracted as $\mathbf{b} = [x_w \ y_w \ l \ age]^T$, where l and age are the ball likelihood and longevity. For a possessed ball, \mathbf{b} is still reported as the possessor's position with $l = 0$. Then, the most likely ball candidate from each camera process is input to the second processing stage, described below in Sections 3, 4 and 5, wherein these observations are combined to estimate the height, phase and trajectory of the motion.

3 Detecting 3D Ball Position and Trajectory

In Section 4 we present a simple yet effective method for recognising the phase of ball motion, which requires cues about the height of the ball. In this Section, two methods are described for detecting that the ball has left the ground plane, and is following a 3D trajectory. The first uses triangulation of multiple sources of data to estimate a 3D position; the second uses an analysis of the trajectory from one (or more) data source to infer that the ball has left the ground.

3.1 Triangulation from Multiple Sources

Assume the ball is observed from two cameras \mathbf{c}_1 and \mathbf{c}_2 with projected positions \mathbf{b}_1 and \mathbf{b}_2 on the ground plane β . We will make a simple estimate \mathbf{x} of the ball position. Let l_1 and l_2 be two lines from \mathbf{c}_1 to \mathbf{b}_1 and \mathbf{c}_2 to \mathbf{b}_2 , respectively. In practice, l_1 and l_2 do not intersect due to errors caused by camera calibration and object detection. Thus, a simple estimate is calculated by assuming errors from both cameras are likely to

be of the same magnitude, and assigning the estimate \mathbf{x} to be the mid-point of the shortest possible line between l_1 and l_2 (see Figure 4).

To find the shortest possible line, two points \mathbf{p}_1 and \mathbf{p}_2 are defined to lie on lines l_1 and l_2 respectively, and it is required that the line $\overline{\mathbf{p}_1\mathbf{p}_2}$ be a common perpendicular of l_1 and l_2 [13]. The points \mathbf{p}_1 and \mathbf{p}_2 can be determined by:

$$(\mathbf{b}_k - \mathbf{c}_k) \times (\mathbf{c}_k - \mathbf{p}_k) = 0 \quad (2)$$

$$(\mathbf{b}_k - \mathbf{c}_k) \cdot (\mathbf{p}_1 - \mathbf{p}_2) = 0 \quad (3)$$

where k ranges over the two line indices $\{1,2\}$. The ball position \mathbf{x} is estimated as the middle point along $\overline{\mathbf{p}_1\mathbf{p}_2}$. If the ball is observed in more than two cameras, we first find the estimated 3D ball position of each pair of different views, and the final ball position \mathbf{x} is estimated as the arithmetic mean of these pair-wise estimates.

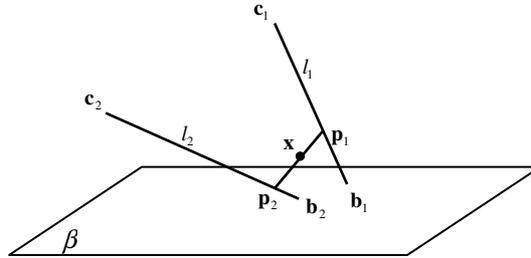


Figure 4: 3D ball estimation using 2D ball positions \mathbf{b}_1 and \mathbf{b}_2 from cameras situated at \mathbf{c}_1 and \mathbf{c}_2

3.2 Recognition of 3D Trajectory

When the ball trajectory is observed by only one camera, cues about its height are extracted from the shape of its trajectory (as opposed to triangulation). We have found two cues to be useful in distinguishing flying from rolling trajectory: first, the detection of the projection of a parabolic (curved) trajectory (as opposed to a straight rolling one); second, the detection of sudden changes of velocity (bounces) in the absence of players.

4 Recognition of Ball Motion Phases

4.1 Phase Definition

The proposal is to classify the ball motion into one of four categories, or *phases*. These are: *rolling* (R), *flying* (F), *in-possession* (P) and *out-of-play* (O). These phases were chosen because they each require different tracking models, though co-incidentally they provide useful insight into the semantic progression of the game. A sequence of play was annotated according to these four definitions, and the graph showing these states is given in Fig 5. In most cases the progression of play is reasonably straightforward to annotate, as a chain of transitions *e.g.* {PIF|O|PIR|P...} between these four phases.

However, there are sometimes ambiguities in interpretation, e.g. between flying and rolling phases or e.g. how many touches of the ball constitutes a possession.

For our purposes, it is useful to denote even a single touch of the ball by a player as a frame of in-possession phase. This is because in-possession phases act as the non-deterministic periods that initialise the rolling and flying phases, literally by kicking them off in a particular direction. Indeed, the ball trajectories for rolling and flying phases are determined by the last kick of a player during the preceding in-possession phase (notwithstanding gusts of wind, and other noise processes). Furthermore, the pattern of play is punctuated by periods when the ball is out-of-play, e.g. caused by fouls, ball crossing touchline, off-side or in-possession by the goal-keeper. Thus, we choose to describe the pattern of play as a list of phase-chains, always started by a in-possession phase, and ending in an out-of-play phase. The phase transition graph for this pattern is drawn in Figure 5.

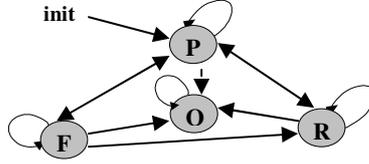


Figure 5: Phase transition graph in soccer ball motion

As shown in Figure 5, a ball is firstly initialized by a player, and then the ball will have phase changes between the four phases. Each phase can last for a certain period of time (frames). When a ball is *out of play*, it will be reinitialized for another cycle of phase transitions (CPT).

Given observations of the ball from separate cameras, and height cues obtained as in Section 3, what follows is the estimation of the current ball phase, given the previous estimates, and the observations output from each camera. In section 5, specific models are described for each of the in-play phases.

4.2 Estimation of Motion Phase

Assume $\mathbf{b}_k = [x_{wk} \ y_{wk} \ l_k \ age_k]^T$ is a 2D ball candidate detected from camera k , and (x_w, y_w, z_w) is the 3D ball position. For each \mathbf{b}_k , $g[\mathbf{b}_k]$ equals to *true* or *false* denoting whether it is between two bouncing points. Then, the ball motion phases are simply determined as follows:

1. A ball is *flying* if $z_w > z_0$ or $\bigcup_k g[\mathbf{b}_k] = true$;
2. A ball is *rolling* if $z_w < z_1$, $\max(l_k) > l_0$ and $\bigcup_k g[\mathbf{b}_k] = false$;
3. A ball is *in-possession* if $z_w < z_1$, $\max(l_k) < l_1$ and $\bigcup_k g[\mathbf{b}_k] = false$;
4. A ball is *out of play* if it cannot be detected in over N_0 frames and the last position reported is near the pitch boundary. An *out of play* ball can be reinitialized through several football events like *throw-in*, *corner-kick* etc.;
5. If not belong to all the cases above, a ball is missed (no detected balls) or uncertain. However, we may still recover their height, as described in section 5.

5 Phase-specific Trajectory Estimates

Finally we describe how the three different models of ball motion are used to provide per-frame estimates of ball position. For *flying* and *in-possession* motion models, the output may lag up to several seconds behind the input observations. This latency is necessary in these cases as to have an estimate at frame $(t+k)$, before publishing an estimate at frame t .

5.1 Estimating the Trajectory of a Rolling Ball

A *rolling* ball can be tracked in the ground plane. We assume that the duration of the *rolling* phase is less than the allowed latency, i.e. all observations of the rolling ball are available prior to generation of the output. At present we model the rolling ball with a simple constant velocity model, i.e. disregarding rolling resistance. It is acknowledged that more complex models would be more appropriate, e.g. with a friction term proportional to velocity. The model parameters are also estimated simply: the start and end positions and times are calculated as the arithmetic mean of all relevant observations at those frames. Again, this estimate can be made more robust by including intermediate observations, weighted by relative uncertainty.

5.2 Estimating the Trajectory of a Flying Ball

The trajectory of a flying ball is modelled as a set of parabolic curves lying in consecutive virtual vertical planes. Each vertical plane is determined by two 3D ball positions estimated in section 3, and in each plane, the trajectory segment is corresponding to a bounce.

To estimate the vertical plane, two 3D ball positions are required, which are extracted in section 3. Suppose π is the determined *virtual vertical plane*, and A is the projected ball position on ground plane β . Given a camera position C , we require a 3D position B . Points C' and B' are vertically below C and B .

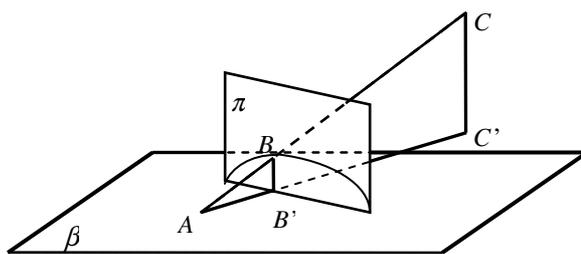


Figure 6: The trajectory is modelled to lie in a vertical plane. Its position in that plane is found by locating the point of intersection between that plane and the line AC

The 3D ball position B is estimated in two stages as follows. First, find B' , defined as the intersection of $\overline{AC'}$ and π . Second, calculate B using the fact that triangles $\triangle ACC'$ and $\triangle ABB'$ are similar, and the latter is known completely. Let us define $X_p = (x_p, y_p, z_p)$ as the world coordinates of any point p , then we have:

$$z_B = \frac{\|X_A - X_B\|}{\|X_A - X_C\|} \cdot z_C \quad (4)$$

For each parabolic segment, three points (lying in the vertical plane) are used to estimate the parameters of this parabola. Similarly to the method proposed for the rolling ball, both model and estimate can be made more sophisticated, however, our aim is the demonstration of an overall system rather than the refinement of one particular component.

5.3 Estimating the Trajectory of an In-Possession Ball

Estimating the trajectory of the football controlled by the player presents two types of difficulty. First, unlike the ballistic phases, the trajectory is a continually stochastic process which is not easily parameterised. Second, the ball is not observed so frequently, as it is often occluded by the player possessing the ball, or another player close by. As a consequence, it is necessary to build in some latency into the estimation process: sightings of the ball are used to back-track its most likely trajectory, during the time in which it was not seen.

In this case, ground-plane trajectories of the players are used to interpolate the most likely trajectory. If the ball was not seen between frames k_1 and k_2 , it was most likely because it was occluded by a player possessing it. The trajectory is estimated to be that of the player closest to the ball position at frame k_1 .

6 Evaluation and Results

We have tested the proposed model in over 5500 frames of a multi-camera sequence. Two types of evaluation are presented. The first type is to compare the estimated and ground-truth (GT) phase estimates, by comparing the recorded transitions, and also the frame-by-frame labels. The second type is to compare the distance (in metres) between estimated and ground-truth ball positions, in which only 2D distance in x-y plane is used. All the ground truth ball positions and phases are derived from manual assistance.

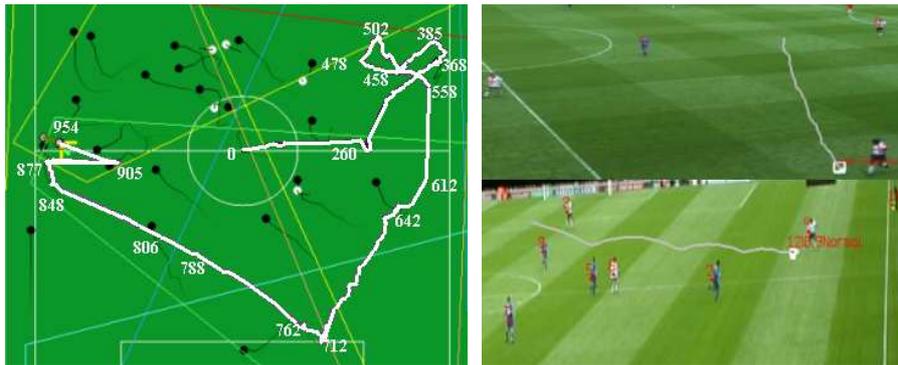


Figure 7: Estimated 3D ball trajectory compared with GT and two 2D trajectories.

Figure 7 gives a complete 3D trajectory history from frame 0 to 954 (left) with two 2D trajectories at frame 831 in camera 2 and 3 (right), respectively. In the 3D trajectory

graph, several frame numbers are marked to denote some of the phase transition points, and player trajectories are also given in black. In eight test sequences of 5500 frames each, we have 3D ball positions estimated in about 3700 frames. Regarding the ball is out of play in 1131 frames, we can recover over 85% in-play balls in our system, with over 90% of them lie within 3m of the ground truth. Considering maximum calibration errors of about 2.5m and inaccurate ground truth for flying balls, our system has very promising performance in real-time ball tracking.

A comparison of phase transitions is shown in Fig 8. It can be seen that all except one of the phase transitions were successfully detected. The missing one phase transition (to a rolling ball) was not detected because there was not a sufficiently clear view of the ball during this motion. The second evaluation measure is the distribution of error distance between the ground truth label and the estimated position, in which about 92% of estimated values lie within 2.5m of the manually recorded position.

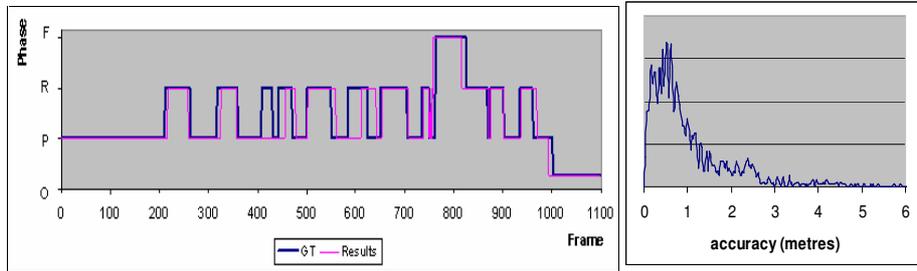


Figure 8: Phase transition graph in a complete CPT with four y-axis positions represent the four phases (Left), and accuracy of our method measured as distance between ground-truth and estimate in the ground plane (Right).

The analysis of the frame-by-frame phases can be presented as a confusion matrix, as in Table 1, from which several facts can be observed. Firstly, during the period, in over 50% of samples, the ball is *in-possession*; and in 33% of samples the ball is *rolling*, thus 2D models can be applied to 83% of cases. Secondly, about 25% *rolling* and 13% *in-possession* balls are misjudged from each other, this happens when a *rolling* ball cannot be observed in a crowd or an *in-possession* ball is rolling near the player who possessed the ball. This affects the accuracy of the ground-truth as well as the estimate from the proposed method. Disregarding the confusion between these two phases, the average correct rate of phase transitions will increase from 82.6% to 98.3%. Moreover, we can also find about 11% *flying* balls are determined as *rolling*. One explanation for this is that a low-flying ball has a similar appearance to a rolling ball. Calculation of the height z is sensitive to errors in camera calibration and motion detection; hence the threshold z_0 has to be tolerant to this error. Heights below z_0 will not be recognized correctly.

GT \ Results	Flying		Rolling		Possessed		Out		Sum	
	frames	%	frames	%	frame	%	frame	%	frame	%
Flying	56	88.9	7	11.1	0	0.0	0	0.0	63	5.7
Rolling	3	0.8	273	73.4	96	25.8	0	0.0	372	33.8
Possessed	1	0.2	77	13.6	480	84.8	8	1.4	566	51.4
Out	0	0.0	0	0.0	0	0.0	100	100	100	9.1
Sum	60	5.4	357	32.4	576	52.3	108	9.8	901	100

Table 1. Quantitative analysis of Figure 8 using ground truth and estimated results.

7 Conclusions

We have presented a model of football motion using distinct phases of trajectory. One interesting feature of the approach is that it uses high-level phase transition information to aid low-level tracking. Through recognition of four phases, phase-specific models are successfully applied in estimating 3D position of the ball. Unlike traditional models, our model can fulfil automatic 3D tracking without shadow information and manual assistance.

The results obtained from our model are very encouraging. Simple mechanisms for classifying the phase of the ball, and estimating its trajectory, are demonstrated to be effective at providing estimate of ball location. There is excellent scope for building more sophisticated models for tracking and content-based understanding of soccer and other videos into this innovative approach.

References

- [1] A. Ekin, M. Tekalp and R. Mehrotra, "Automatic soccer video analysis and summarization", *IEEE Trans. on Image Processing*, 12(7): 796-807, 2003.
- [2] T. Bebie and H. Bieri, "SoccerMan – reconstructing soccer game from video sequence", *Proc. ICIP*, 898-902, 2000.
- [3] K. Matsumoto, S. Sudo, H. Saito and S. Ozawa, "Optimized camera viewpoint determination system for soccer game broadcasting", *IAPR Workshop on Machine Vision Apps*, 115-118, 2000.
- [4] Y. Seo, S. Choi, H. Kim, K. S. Hong, "Where are the ball and players?: soccer game analysis with color based tracking and image mosaick", *Proc. ICIAP*, 196-203, 1997.
- [5] T. D’Orazio, C. Guaragnella, M. Leo and A. Distanto, "A new algorithm for ball recognition using circle Hough transform and neural classifier", *Pattern Recognition*, 37: 393-408, 2004.
- [6] T. Kim, Y. Seo and K. S. Hong, "Physics-based 3D position analysis of a soccer ball from monocular image sequences", *Proc. ICCV*, 721-726, 1998.
- [7] Y. Ohno, J. Miura and Y. Shirai, "Tracking players and estimation of the 3D position of a ball in soccer games", *Proc. ICPR*, 145-148, 2000.
- [8] V. Tovinkere, R. J. Qian, "Detecting semantic events in soccer games: toward a complete solution", *Proc. ICME*, 1040-1043, 2001.
- [9] M. Xu and T. Ellis, "Partial observation vs. blind tracking through occlusion", *Proc. BMVC*, 777-786, 2002.
- [10] N. Babaguchi, Y. Kawai and T. Kitashi, "Event based indexing of broadcasted sports video by intermodal collaboration", *IEEE Trans. Multimedia*, 4: 68-75, 2002.
- [11] I. Reid and A. North, "3D trajectories from a single viewpoint using shadows", *Proc. BMVC*, 863-872, 1998.
- [12] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking", *Proc. CVPR*, 246-252, 1999.
- [13] J. Ren, J. Orwell, G. A. Jones and M. Xu, "A novel framework for 3D soccer ball estimation and tracking", *Proc. ICIP*, 2004.