Estimating the Position of a Football from Multiple Image Sequences

Jinchang Ren, James Orwell and Graeme A. Jones
Digital Imaging Research Center, Kingston University
Penrhyn Road, Kingston upon Thames, Surrey, KT1 2EE, UK
{j.ren, j.orwell, g.jones}@king.ac.uk

1 Introduction

This paper describes research into techniques to estimate the position of the ball during a football (soccer) match. The source data is eight video sequences captured from stationary digital cameras: every point on the pitch is included in at least one camera. However, the ball in play is not always observed in at least one camera: at times it is occluded by players, and other times outside the field of view, a consequence of being kicked high in the air. The focus of this paper is how the position of the ball can be estimated in these cases, by using data from both before and after the event.

In a true real-time system, the estimate is required instantly; implying only prior information about the ball can be used. In an offline version, all prior and subsequent information can be employed in making the estimate of ball position. This paper addresses the border between these two paradigms: a system that is real-time with finite latency. Thus, the problem is cast as estimating the ball position, given all prior frames in the sequences, and *n* subsequent frames. The estimate is therefore subject to an *n*-frame delay, excluding processing time. The architecture designed to implement this system is described in Section 3. In Section 4, we briefly describe the method for extracting moving objects and assigning each a likelihood of representing the ball.

The key constraint to be exploited is that there is exactly one ball in the scene (during times when the ball is in play). This constraint is used to determine the path of the ball in two challenging cases, discussed in Section 5: when it is outside all fields of view (Section 6); or persistently occluded by a player (Section 7).

2 Previous work

Tracking the positions of football players and ball has been attempted from both broadcast television data (moving cameras) [1,2,6,8] and also fixed cameras [3,4,5,7]. Various methods are proposed to detect observations of the ball, including template [5] and colour [6] matching, Hough Transform [7] and, for fixed camera input, background subtraction. For fixed camera work, partial coverage is obtained using two [3] and four [5] cameras, complete coverage requires eight cameras [4]. 3D models of a ball are presented in [3] and [4], though without the accompanying ball detection process. Mining the trajectory of the football has been introduced to in [8].

3 System description

The input data is PAL size colour frames at 25 fps from 8 static cameras distributed around the stadium. The data from each camera is processed to obtain descriptions of the objects moving within that field of view. Each description includes a measure of the likelihood that the object represents the ball. All object descriptions, from all cameras, are sent to a subsequent tracker process, to estimate the ball position. Camera calibration information is available to use in both detection and tracking processes.

4 Detection of ball-like features

The moving objects are detected using foreground-background differencing. A connected components analysis segments the foreground data into distinct objects. Each object is tracked with a Kalman filter, providing an estimate of object speed and allowing some splitting of merged objects [9]. All sizes and speeds can be expressed in world co-ordinates, using the camera calibration data available, and assuming the object is on the ground plane. (This assumption is only violated when the ball is airborne). A heuristic method, incorporating the size, shape, speed and colour of the object, is used to assign to each object a measure, from 0 to 1, of the likelihood that it represents the ball. The speed of the ball was found to be an important cue. Not only is the ball is usually the fastest object on the pitch, but its image will include a motion-blur proportional to the speed. The size filter should account for this effect, and also the error in size calculation caused when the ball is flying (and therefore does not lie in the ground-plane).

The output of this process is usefully visualized in two dimensions by the type of plot shown in Figure 1.

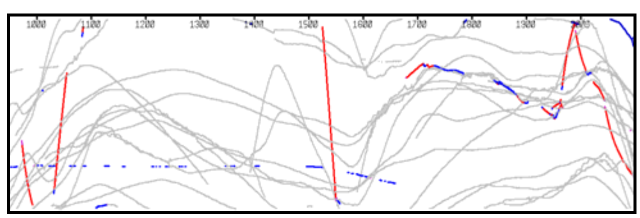

Figure 1: Spatiotemporal plot of features for a single camera. Grey features are the players; blue and red are estimated to represent the ball with low and high likelihood respectively.

Here, the frame number (time) is plotted along the x axis, and the y axis represents a spatial co-ordinate – in this case, the *x* co-ordinate of the image plane. Thus, a ball moving from left to right across the image is represented by a downward-sloping line.

5 Classification of the ball state

During the course of a game of football, the ball is subject to various modes of motion, *e.g.* being handled, kicked, thrown, dribbled and trapped in the netting of the goal. A simplified taxonomy of motion is used in this work: the ball is classified as *rolling*, *flying* or *possessed*. The rolling and flying states describe a ball that has been kicked along the ground, and into the air, respectively.

For a *rolling* ball, the vectors describing motion and position lie in the ground plane, and can be simply estimated from a single camera. The extra degree of freedom enjoyed by a *flying* ball implies its 3D position cannot be estimated from an instantaneous estimate from a single camera. This is discussed in Section 6.

Moreover, a ball *possessed* by a player is likely to generate few if any observations, as the player and ball often form a single connected component, not recognized as representing a ball. In this case, the proposed strategy is identify the player possessing the ball, and to estimate the ball position at the feet of this player during this state. This is described in Section 7.

6 Estimation of 3D position

Two types of strategy are used to estimate the position of the flying ball, depending on how many cameras provide observations. If observations are available from more than one camera, then triangulation is possible.

If only one camera source is available, the sequence of observations can be used to constrain the parameters of allowed trajectory models. The bounce of a ball is an important cue. Assuming a simple model of the bouncing ball, the detection of two consecutive bounces (possibly each from a separate camera) allows estimation of the vertical plane in which the ball in bouncing. In this context, the initial kick that caused the flight may be regarded as the first bounce. Estimates of the 3D positions of the ball between two bounces require the position of the second bounce, and can therefore only be output after this has been estimated. Thus, a latency of around 2-3 seconds is required by the algorithm. One method for bounce detection is to threshold the curvature on the ball trajectory; results on this will be shown.

7 Estimating position of the ball in possession

There are extended periods of time in which the ball in play is not able to be directly observed by the system – as it is occluded by players in close proximity. Nevertheless, we still require an estimate of ball position during these times. The proposed procedure is as follows. We assume a direct observation is available immediately before the start of the occlusion. A possession hypothesis is then generated for each player in that local vicinity at the time. Often, this is just one player; but for ambiguous occlusions the following method is proposed. Once a subsequent direct observation is available (after this occlusion), the closest player to this observation is assumed to have been in possession of the ball. The ball position is then estimated to be at this player's feet, over the duration of the occlusion. This 'Track-Back' technique is only possible if the subsequent observation is available before the occlusion estimate is required, i.e. in an offline system, or the latency is greater than the duration of the ambiguous occlusion.

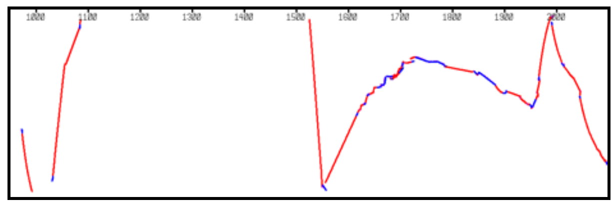

Figure 2: The same spatiotemporal plot as Fig. 1, showing only the path of the ball, using Track-Back technique to interpolate over occluded data.

8 Conclusion

A method for estimating the position of a football from multiple static sequences is proposed. Three modes are used to model the ball state, and techniques are outlined for each of these. For the airborne and possessed football, analysis of the spatiotemporal sequence of observations is necessary to overcome inherent ambiguities.

References

[1] Y. Gong etc., "Automatic Parsing of TV Soccer Programs", in Proc. Multimedia Comp. and Systems, 1995, pp. 167-174.

[2] D. Yow etc., "Analysis and Presentation of Soccer Highlights from Digital Video", ACCV'95, pp. 499-503.

[3] T. Bebie, and H. Bieri, "SoccerMan – Reconstructing Soccer Game from Video Sequence", in ICIP, 1998, pp. 898-902.

[4] Y. Ohno etc., "Tracking Players and Estimation of the 3D Position of a Ball in Soccer Games", ICPR'00, pp. 145-148.

[5] K. Matsumoto etc. "Optimized Camera Viewpoint Determination System for Soccer Game Broadcasting", IAPR Workshop MVA, 2000, pp. 115-118.

[6] Y. Seo etc., "Where Are the Ball and Players?: Soccer Game Analysis with Color Based Tracking and Image Mosaick", in ICIAP, 1997, pp. 196-203.

[7] T. D'Orazio etc., "A Ball Detection Algorithm for Real Soccer Image Sequences", in ICPR, 2002, pp 201-213.

[8] X. Yu etc., "Trajectory-Based Ball Detection and Tracking with Applications to Semantic Analysis of Broadcast Soccer Video", in Proc. ACM Multimedia, 2003, pp. 11-20.

[9] M. Xu and T. J. Ellis, "Partial Observation vs. Blind Tracking through Occlusion", in Proc. BMVC, Cardiff, Sept 2002, pp. 777-786.