HAWK-EYE TENNIS SYSTEM

N. Owens, C Harris, C Stennett

Roke Manor Research, United Kingdom

Abstract

A description of the Hawk-Eye tennis ball tracking system is presented, detailing the major design considerations, and how the various technical challenges were overcome. This system is used in the sports broadcasting area.

1. INTRODUCTION

Television broadcasters are continually looking for ways to enhance their coverage of sport. The rationale for developing the Hawk-Eye Tennis System was to provide a low-cost but easily deployable passive system for tracking the tennis ball during a match. The resulting information can then be used in virtual reality replays and to provide statistical information for commentators. Hence, the system had to be near real-time, providing a viewable track within 5 seconds of the end of the rally.

Tracking objects in professional sporting events differs from most tracking problems due to the requirement for a completely passive system. Vision systems provide an appealing answer due to their inherently passive nature. Other solutions such as radio transponders are unattractive due to the requirement to tamper with the ball or the players.

Ease of installation also provides a major desirable feature. Typically TV production companies rig and de-rig grounds in a day, and doubling the number of cameras can require many kilometres of extra cable for the company to rig, at significant labour and material cost. The system had to make use of the existing broadcast cameras, which were panning, tilting and zooming outside of the operators' control.

2. SYSTEM OVERVIEW

A decision was made early on in the development to perform 2D tracking on each camera, and combine the 2D tracks elsewhere. With respect to the standard technique, namely reconstitution on a per-ball basis, this had three advantages: (a) providing a convenient computational division of labour within the system; (b) reducing the combinatorial comparisons required; (c) reducing the required comms bandwidth.

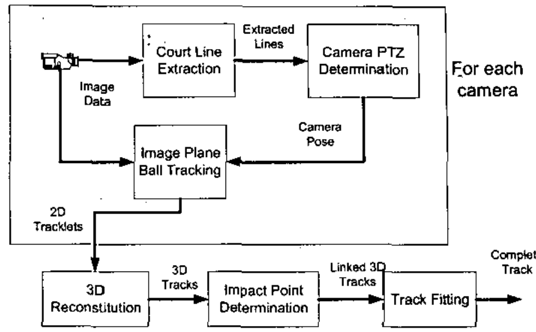

Figure 1: System Block Diagram

For each frame on each camera, the camera PTZ position is determined by tracking the court lines. The ball is extracted on each frame and 2D 'tracklets' of the ball's motion are built up on the image plane, using the frame's associated extrinsic calibration to correct for camera motion. These tracklets are sent to the 3D Reconstitution module which constructs the tracklets into 3D tracks, and determines the impact points between separate tracks (can occur at a bounce or a strike). The completed track is then sent to the visualisation software for display and analysis.

3. REAL TIME ISSUES

A major requirement of the system was the ability to work at real-time. The completed track was required on screen no more than 5 seconds after the end of a point. To provide the necessary computing power, a separate PC dealt with each camera. The 3D Reconstitution and tracking was performed on another PC, as was the visualisation. Data was transmitted from platform to platform over TCP/IP.

4. CAMERA CALIBRATION AND PTZ DETERMINATION

For each camera, the intrinsic calibration (the mapping of image pixels to and from spatial directions), and the pan/tilt/zoom (PTZ) must be determined (see below). These two tasks have much in common, and are performed by use of model-based tracking, [Harris and Stennett (1), Harris (2)].

Model-based tracking determines the system parameters that provide registration between a known 3D model and corresponding image observations. For geometric camera calibration, the parameters are both the intrinsic camera parameters, and the camera location and attitude. For PTZ, only the camera attitude and focal length are sought, as the camera location and other intrinsic parameters remain fixed.

The 3D model needs to be both accurately known and visually prominent - and fortunately the lines of the tennis court satisfy these requirements. Since the apparent width of the imaged lines can vary due to imaging effects (such as gamma correction and choice of threshold), only the line centres are used. A series of control points are positioned along the model court lines. The image processing starts by extracting (nominally) straight lines by use of spatially adaptive thresholding (Local Mean Removal), and retaining long segments after straight-line decomposition (this provides robustness against non-modelled image features, such as the players). Each control point is satisfied by finding a pair of segments (one from either side of the court lines) close to the requisite angle, and their mean location taken. The mis-registration of the satisfied control points is used to perform iterative parametric optimisation by using the appropriate parameter differentials.

5. CAMERA CALIBRATION

This is the determination of the intrinsic geometric mapping of image pixels to and from spatial directions. A parametric form is used, comprising the focal length, aspect ratio and location of the optical centre (where the optical axis strikes the image plane). There are radial distortions (e.g. barrel distortion), which need to be determined (a single term was found to be sufficient), and, as these vary with zoom, these have to determined across the zoom range. In practice the calibration is undertaken at a number of zooms, and piecewise linear approximation used to transition between the recorded points [Harris (4)].

It is important that camera calibration is performed most accurately, as errors introduced here will propagate. Because of this, calibration is undertaken on a full resolution static image (avoiding field tear permits double resolution), and with the court fully filling the field of view. The camera location is also obtained from the camera calibration.

6. CAMERA PTZ DETERMINATION

Broadcast cameras are usually undergoing continuous pan/tilt/zoom (PTZ) motion about a fixed location. This motion must be measured on each frame before any subsequently extracted ball measurements will make sense. The model-based tracking is used to solve only for the 4 parameters of camera attitude and focal length. As the frame-to-frame camera motion may be large, a simple alpha-beta predictor is used to initiate the parameters on each new frame.

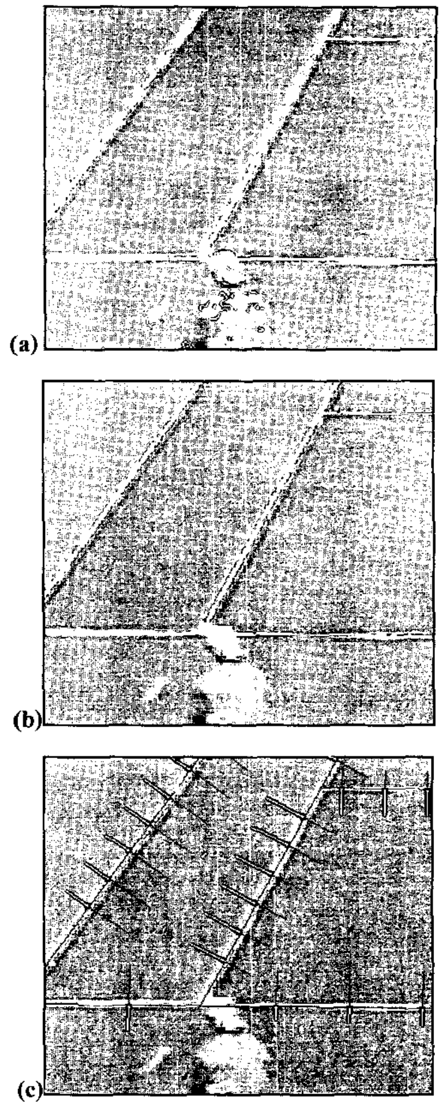

Figure 2: Court Line Extraction. (a) LMR + subpixel boundary walk produced image loops; (b) Straight line portions of loops; (c) Straight line portions mapped back onto court model with centre lines marked. Perpendicular lines show magnified residual to each line edge.

Tracking may on occasion be lost, when an insufficient number of the court lines are seen - for example, when the camera may follow a high lob, or zooms in on a serve. Re-acquisition of the court tracking must be performed automatically. This is performed by conducting a binary subdivision search across the PTZ parameters, and terminating when successful tracking is achieved. To ensure that a hopeless image is not dwelled upon, the search is performed in ~1/25th second portions on the most recent frame.

7. IMAGE PLANE BALL TRACKING

In the imagery, a tennis ball typically has a diameter of some 2 to 10 pixels, depending on zoom and range. If the ball is moving slowly in the image, or the camera has a fast shutter, it appears as a circle (or ellipse if the pixel aspect ratio departs from unity). However, broadcast cameras often have slow shutters, so that significant motion blur can occur, and in them the ball appears as a sausage-shaped streak.

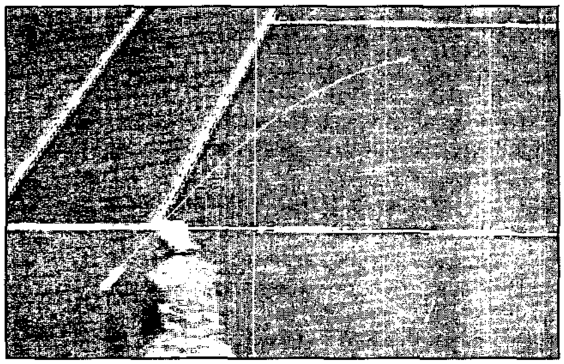

Figure 3: Ball Extraction. The ball track has been remapped into the current frame's calibration. The ball is the sausage shaped object to the left of the player.

Ball extraction is performed using the LMR algorithm again, which results in a number of closed boundaries (loops), originating from the ball, players, court lines, net, etc. These are filtered on size and shape to give candidate balls. Using the camera calibration and PTZ, their centroids are transformed into the coordinates of a nominal stationary camera, and in these coordinates ball tracks are formed. It is vital that the tracks are constructed in a stationary camera, because the raw image coordinates may be dominated by the effect of PTZ motion.

A valid ball track will be smooth and down-curving. The discontinuities that occur when the ball bounces or is struck will result in separate tracks. Tracks are initiated from all unmatched candidates, and are continued using a polynomial fit to predict the ball's location in subsequent images. The polynomial fitting order increases according to a pre-defined program as the track length increases, to allow principally for the effects of perspective projection. Similarly, the radius of the acceptance region for matching varies according to the duration over which matching has failed to occur. Matching criteria include position, shape consistency, number of candidates, and, for long shutter times, the ball streak vector. When tracking fails for too great an interval, the track is deemed complete. Completed tracks are assessed for trajectory quality (mainly length and movement), and acceptable ones are passed onwards for 3D association and interpretation. Although this results in a number of spurious tracks, it avoids making a (rash) early decision on which is the true ball track.

8. 3D RECONSTITUTION

The 3D Reconstitution module recombines the 2D tracklets into 3D tracks, finds impact point locations between the 3D tracks, and does curve fitting to provide a smooth completed track.

Not much filtering of the tracklets is done at the image processing stage, since when working in 3D the epipolar constraint is powerful as a noise filter. Hence this module must deal with a lot of spurious tracklets.

Firstly, the module examines pairs of tracklets that overlap in time, and attempts to match them up into Partial Tracks, short 3D portions of a track. The incoming tracklets are sorted into time bins, making the initial matching faster. An initial test is used to test a pair of tracklets for matching. Three equally spaced points within the overlap region are triangulated, and if two out of the three points produce acceptable image plane error, the triangulated points over the whole overlap region are computed, and if the mean squared pixel error between the candidate Partial Track and original tracklets is small enough, the candidate is added to the list. To reduce cabling, the cameras were not synchronised so a linear interpolation between the second tracklet's data points is done to match a reading from one tracklet to another.

These Partial Tracks are modelled in 3D using a simpler polynomial model, and are then matched up together into collections of Partial Tracks called Compound Tracks. A compound track represents the flight of the ball between two impact points (against the racket, ground or net).

The real-time nature of this system and high number of spurious tracklets demanded a solution that avoided the large combinatorics of hypothesis type systems. At each stage, the data represented high and higher level elements of the ball flight. Each track is represented by a number of 2D tracklets, perhaps 20 tracklets for a long track. Those tracklets are then triangulated to a number of 3D partial tracks, perhaps 3 or 4. Those partial tracks are then combined into a single compound track, representing the flight of the ball between two impact points.

9. IMPACT POINT DETERMINATION

This module then attempts to join up compound tracks by finding common impact points between them. A quadratic model is used to find an approximate join point between two compound tracks. This estimate is then used to initialise a Kalman Filter [Kalman (4)]. This Kalman Filter takes in the 2D tracklet data directly, and iteratively tries a linear, quadratic and cubic model of the incoming and outgoing compound track to get the best estimate of the impact point possible. A rule-based system is used to decide whether an impact is a bounce, a strike or a half volley, using the velocity directions and positions to reduce the degrees of freedom. A simple distance metric is used to determine which impact point match is the most likely.

Special code was written to deal with the case of a half volley. Often the upward bouncing portion of the half volley was too short to be extracted, and so an estimated 'dummy' track was inserted in the impact point chain. The bounce point was determined using a one-sided Kalman Filter applied to the incoming track. Configurable bounce retardation percentages were used to estimate the initial velocity of the upward bouncing track at the instant just after the bounce. This was used to create a purely ballistic intermediate track, with an acceleration $\mathbf{a}$ of $[0, 0, -g]^T$. This dummy track was inserted in the chain of tracks, providing an unbroken set of tracks for the visualisation module.

10. TRACK FITTING

Finally each completed track is fitted to a polynomial, using the criteria that the 3D curve must pass through the calculated impact points, but minimise the mean squared error from the short partial tracks. The degrees of freedom in the polynomial model are a function of the track length, but typically the function is either a quartic or a quintic. The model parameters for each portion of the track are then sent to the visualisation module.

11. VISUALISATION

The visualisation module allowed an operator to show the movement of the ball on a virtual court. The application used the Direct3D interface in Microsoft's DirectX API to render the graphics. The operator was able to move around in time and space in the virtual world to show whatever aspect of the rally was most interesting and pertinent at that time. The instantaneous speed of the ball was also shown to give the commentators information on the speed of serve, ball slowdown out of a bounce, return speeds, etc.

12. PERFORMANCE

Television tools have a very low false alarm tolerance, and a single televised mistake can have a huge impact on the systems credibility. In this context the continued use of the system points to a good level of performance.

Several diagnostic tools and displays are built into the system, allowing the operators to quickly assess tracking quality and filter out poor tracks before they are made available for broadcast. We have no quantified assessments but we believe that about 90% of requested sequences are of sufficient quality to broadcast.

13. CONCLUSIONS

- Modularisation improved the speed of system development. Splitting up the image processing and the 3D reconstruction provided the most efficient design development, allowing the system to go from requirement to deployment at the Davis Cup in the UK in 12 weeks.

- System performance has proved very good, and the completed track was available at the visualisation software 3-4 seconds after the end of the rally. After initial use at the Davis Cup in the UK, Hawk-Eye tennis has since been deployed at the Masters' Tournament in the United States and at the Queens tournament.

14. ACKNOWLEDGEMENTS

The development of the Hawk-Eye Tennis system has been funded by Hawk-Eye Innovations Ltd, The Hampshire Rose Bowl, Southampton, UK.

15. REFERENCES

1. C G Harris and C Stennett , 1990, "RAPID - A video rate object tracker", Proc. of the BMVC, 1990, Oxford, 73

2. C G Harris, 1992, "Active Vision", edited by A Blake, A, Yuille, The MIT Press, "Tracking with Rigid Models", 59-73

3. R E Kalman, 1960, "A new approach to linear filtering and prediction problems", ASME J. of Basic Engineering.

4. C G Harris, 1993, "Geometric Camera Calibration for Vision Based Navigation", Intelligent Autonomous Vehicles, 1993, 77-82