Abstract

This paper presents a real-time computer vision system that tracks the motion of a tennis ball in 3D using multiple cameras. Ball tracking enables virtual replays, new game statistics, and other visualizations which result in very new ways of experiencing and analyzing tennis matches. The system has been used in international television broadcasts and webcasts of more than 15 matches. Six cameras around a stadium, divided into four pairs, are currently used to track the ball on serves which sometimes exceed speeds of 225 kmph. A multi-threaded approach is taken to tracking where each thread tracks the ball in a pair of cameras based on motion, intensity and shape, performs stereo matching to obtain the 3D trajectory, detects when a ball goes out of view of its camera pair, and initializes and triggers a subsequent thread. This efficient approach is scalable to many more cameras tracking multiple objects. The ready acceptance of the system indicates the growing potential for multi-camera based real-time tracking in broadcast applications.

1. Introduction

Computer vision has rich potential for enabling a new class of applications involving broadcasting, viewing of remote events and telepresence. While the research community has recognized the importance of video analysis and multi-camera based systems for interactivity, immersion, and search (see, for instance [5, 4, 1, 3, 6]), there have been few reports of successful deployment of such systems in real-world applications. Over the last two years, we have developed and deployed a multi-camera based real-time tracking and visualization system which is being regarded as introducing a new paradigm into sports broadcasting.

The system tracks the players and the ball in tennis matches in real time to introduce a number of innovations in live television broadcasts and webcasts of international tournaments. While the player tracking system has been described before [2], this paper focuses on the recently introduced multi-camera ball tracking system. Ball tracking is challenging because of the small size of the ball (67 mm in diameter), the relatively long distances it travels (over 26 m in length), the high speeds at which it travels (the fastest serves are over 225 kmph), changing lighting conditions, especially in outdoor events, and varying contrast between the ball and the background across the scene.

The configuration of the ball tracking system and the tracking approach are described in section 2. Section 3 presents new visualizations enabled by ball tracking which have been readily accepted and extensively used by broadcasters.

2. Ball Tracking

2.1. System design and configuration

The system consists of six monochrome progressive scan (60 Hz) cameras connected to a quad-pentium workstation with dual PCI bus. The number of cameras was chosen to be sufficient to cover the volume over which the ball typically travels and to ensure that the field of view of each camera is not too large. Progressive scan cameras which operate at 60 Hz are needed to capture images with temporal resolution good enough for ball tracking and so that the ball appears big enough in the image to be tracked. While cameras with an even higher speed and resolution could be used, we have chosen these based on the constraint of real-time processing and to stay within cost bounds. A general purpose computer has been used to ensure ease in developing the system, integrating it with other systems, and upgrading. Monochrome, rather than color cameras, have been chosen so that the bandwidth of a dual PCI bus is sufficient for full-frame capture at 60 Hz from all six cameras. Thus, color which is a strong cue for ball tracking, had to be sacrificed to meet the other system constraints. This has made ball segmentation a more challenging problem.
The six cameras are placed around a stadium with four cameras on the side and two at the ends of the court. Each of the four side cameras is paired with one of the end cameras to form a set of four stereo pairs that track the ball in 3D. Auto-iris lenses are used with the cameras to cope with large changes in lighting in the course of a day. Additionally, tracking parameters are dynamically updated, as explained in section 2.3.

2.2. Multi-threaded tracking

A multi-threaded approach is taken to tracking to achieve an efficient (real-time) solution that is scalable and works with distributed computing resources. A processing thread is associated with each camera pair. Figure 1 gives an overview of the processing steps in each thread. Each thread waits for a trigger signal to start frame capture and processing. Each thread has the following set of parameters: a trigger to start processing, a pair of associated cameras, calibration parameters of each camera, difference image thresholds for each camera, ball size parameters, expected intensity range for the ball, expected ball position in each camera image, size of the search window in each camera image, a trigger signal for the subsequent processing thread, and a pointer to the parameters of the subsequent thread. Camera calibration parameters are obtained beforehand using the algorithm in [7], taking advantage of the calibration grid provided by the court itself.

On receiving its trigger, a thread executes a loop of capturing frames from the camera pair, detecting the ball in the captured frames, stereo matching, and updating the 3D trajectory and tracking parameters, until the ball goes out of view of any one of its associated cameras. At this time, the current thread initializes the parameters for the thread corresponding to the subsequent camera pair and triggers that thread.

This multi-threaded approach scales in a straightforward manner to any number of cameras tracking an object over a large area. With a few modifications, the approach also scales to multiple objects being tracked by multiple cameras. In this case, a thread associated with a camera pair (or set of cameras) has triggers associated with each object to be tracked. The thread tracks an object when it receives a trigger signal corresponding to the object. Different tracking schemes can be used by a thread for different types of objects. For instance, different sets of parameters and tracking schemes are used for tracking the players and the ball in the tennis application.

2.3. Ball segmentation and detection

Ball segmentation/detection in any camera image is achieved by frame differencing the current and previous images and thresholding the result, finding the regions in the current image that lie in the expected intensity range for the ball, performing a logical AND operation of the regions obtained from the preceding two steps, subjecting the resulting regions to size and shape (aspect ratio) checks, and choosing the detection closest to the expected position in the (rare) case of multiple detections. All these operations are performed only in a window defined by the expected ball position and search size parameters. Most parameters, such as the range of intensity values, expected size, size of the search window, and the differencing threshold, are dynamically updated during the course of tracking. The expected position of the ball is continually updated based on the current velocity of the ball.

Parameters such as the search size and range of intensity values are initially set to conservative values. The first thread to be triggered is determined based on the direction of the serve. This thread initially has no expected ball position but a relatively large search window. The search for the ball is performed in only one of the two cameras to ensure efficiency. Once the ball is detected in one camera, the search region in the other camera is based on epipolar constraints. Once tracking commences, the search regions are tightly defined and ball detection is performed in parallel in images from both cameras. When the current velocity of the ball indicates that the ball will be out of bounds of the current camera pair by the next frame, the positions of

Figure 1. Overview of processing in each thread associated with a camera pair
the ball in the next camera pair are determined based on the current 3D velocity of the ball and 3D to image mapping using camera calibration parameters. Thus, once the initial thread starts tracking, subsequent threads look for the ball in well-defined search windows. The dynamic update of segmentation and tracking parameters has been key to the success of this system.

2.4. Landing spot determination

The ball landing spot for each serve is determined by analyzing the 3D trajectory obtained from tracking and performing an appropriate interpolation. Interpolation is performed where the z (height) component of the velocity changes from negative to positive, indicating the ball stops falling and begins rising. If the 3D trajectory of length \( n \) has time samples \( (t_1, t_2, \ldots, t_n) \), and the time sample \( t_c \) represents the last sample with a negative z velocity (computed from time \( t_{c-1} \) to \( t_c \)), then the landing spot is at a time \( t_l \) which is either between \( t_c \) and \( t_{c+1} \) or between \( t_{c-1} \) and \( t_c \).

In the first case, the 3D velocity and acceleration parameters at time \( t_c \) are projected forward to determine when the ball would reach the ground. In the second case, the velocity and acceleration parameters at time \( t_{c+1} \) are projected backwards to determine the landing spot and landing time. The choice between the two is made based on how well the new velocity resulting from the interpolation blends with the known velocities. Our experiments have shown that the choice is unambiguous.

![Figure 2. Example of tracking the ball in a pair of cameras](image)

2.5. Results and Accuracy Issues

Ball tracking has been successfully performed on hundreds of serves in more than 15 matches at the 1999 World Championship in Hannover and the 1999 Paris Open. Figure 2 shows an example of ball tracking in one pair of cameras. We have achieved real-time tracking and have verified that the hand-off between cameras or threads is smooth. The 3D trajectories are stored in a database along with a host of other information related to the match as explained in section 3.1. The visualizations and statistics obtained from the ball trajectories have been eagerly used in numerous international live television broadcasts and webcasts. Some of the results are presented in section 3.

We have verified the accuracy of the system by several means: a) determining the accuracy of ball tracking in the image plane by careful analysis on a number of recorded sequences – these have shown ball tracking accuracy to be within a pixel; b) determining the accuracy of image to 3D mapping based on test points – these have shown object space errors under 15mm; c) careful comparison of trajectories and landing spots obtained by tracking with video sequences from different broadcast cameras; and d) comparison of speed of serve at racket obtained from ball tracking with that obtained by a radar gun – the difference has been within 10 kmph and has been partly due to a lack of precise match in the times at which the two speeds are measured. All these experiments indicate positional accuracy to be within a few millimeters and typically within 20 mm. However, it would be useful to further determine accuracy using an independent high speed modality for capturing ground truth data.

3. Virtual Replays and Visualizations

Ball tracking has enabled an exciting set of visualizations of a tennis match, immersive for a viewer. Users decide which part of the game is interesting to them and which visualization they wish to explore.

3.1. Data selection

The database contains a description of both dynamic and static aspects of the real-world environment. Ball trajectories are one example of dynamic data, together with player motion trajectories [2] and the changing game score. Tennis court geometry and player and tournament related information are examples of static data.

A powerful selection mechanism allows the user to choose any subset of this voluminous data. In the context of tennis, this selection includes score-based queries (e.g. All serves won by a player), statistics-based queries (e.g. All serves at a speed above 200kmph) or space-based queries (e.g. All serves directed to the right corner of the left service box). Each query can be further narrowed down using a time constraint, for example limiting it to one set, one game, or even any particular match period (e.g. All serves in the first half hour of the match). In addition, the system supports historical queries, allowing data selection across several matches (e.g. All aces by Sampras against Agassi in 1999).
3.2. Virtual mixing console

Having selected a data subset, the user is presented with a set of tools for viewing and analysis. The concept of a virtual mixing console is used to facilitate visualization selection, smooth transition between different visualizations, and combination of several visualizations. Selected visualizations share space in a visualization window. Currently, three types of visualizations are offered: virtual replays, service landing positions, and serve statistics, described in more detail in the following sections. A new type of visualization can be easily plugged into this scheme.

3.3. Virtual replays

Any particular serve can be viewed from any point of view at any speed. For instance, a spectator can become the receiver of a serve and appreciate the dynamics of the game from this position. A large variety of statistics can be offered for each serve, including the ball’s speed at the racket, its speed after bounce, and the height at which it passed above the net or at which it reached the receiver.

Figure 3 shows a match-point ace, served by Sampras in the final against Agassi during the 1999 ATP World Championships. Figure 4 presents the serve by Agassi which reached the greatest height at the receiver, in the same match. Comparing even these two examples shows how wide a variety of serves one can expect in a tennis match.

3.4. Multiple serve visualizations

A sequence of serves can be selected and shown as consecutive virtual replays to reveal the serving style of a player or to compare the styles of different players. Figure 5 shows superimposed trajectories from virtual replays of all aces served by Pete Sampras during the final of 1999 ATP World Championships against Andre Agassi. The density of aces

Another way of analyzing multiple serves is a Service landing position map which gathers spots on the court where the players direct their serves. Figure 6 shows landing position maps for the 1999 Paris Open finalists, Andre Agassi and Marat Safin. Agassi serves very precisely and consistently into the corners of the serving box while Safin’s serves are more spread out with all his second serves going into the center.

If a more detailed query is used for data selection, the service patterns can be analyzed in more detail. For instance, Figure 7 shows two maps for Safin, showing his serves won and lost. A careful viewer will notice that he often lost points when serving into the center of the serving box and far from the service line.
4. Conclusions

This paper presented a multi-camera real-time ball tracking system that has been used in live broadcasts by more than 20 television networks broadcasting in more than 70 countries, in addition to webcasts from an official tennis website (atptour.com). The approach taken here exemplifies several aspects of designing a successful visual tracking system. These include: a judicious choice of cameras – their number, placement, type, and speed; use of tight constraints provided by the environment and the application; highly efficient processing for real-time operation; dynamic update of parameters to cope with environmental changes; scalability of the solution to increased number of tracked objects, increase in covered area, increased number of cameras, and increased computational and bandwidth resources; and integration of tracking with a larger system to make it relevant to the application at hand.

Future work includes dealing with player-ball interactions and occlusion of the ball due to players, tracking the ball through out a game rather than just on serves, extension of the multi-threaded tracking approach to other applications, and experiments for better determination of accuracy using an independent high speed modality as ground truth.

References


