Visualization of Sports using Motion Trajectories: Providing Insights into Performance, Style, and Strategy

Gopal Pingali\(^1\)* Agata Opalach\(^2\) Yves Jean\(^3\)* Ingrid Carlberg\(^3\)

gpingali@us.ibm.com agata@lucent.com yvesjean@avaya.com carlbom@lucent.com

1 IBM T.J. Watson Research Center
30 Sawmill River Road
Hawthorne, NY 10532, USA

2 Bell Labs, Lucent Technologies
600 Mountain Avenue
Murray Hill, NJ 07974, USA

3 Avaya Labs Research
666 Fifth Avenue, 11th Floor
New York, NY 10103, USA

Abstract

Remote experience of sporting events has thus far been limited mostly to watching video and the scores and statistics associated with the sport. However, a fast-developing trend is the use of visualization techniques to give new insights into performance, style, and strategy of the players. Automated techniques can extract accurate information from video about player performance that not even the most skilled observer is able to discern. When presented as static images or as a three-dimensional virtual replay, this information makes viewing a game an entirely new and exciting experience.

This paper presents one such sports visualization system called LucentVision, which has been developed for the sport of tennis. LucentVision uses real-time video analysis to obtain motion trajectories of players and the ball, and offers a rich set of visualization options based on this trajectory data. The system has been used extensively in the broadcast of international tennis tournaments, both on television and the Internet.

Keywords: sports visualization, virtual environment, telepresence, real-time video analysis, multi-camera tracking, multimedia indexing.

1 Introduction

Sporting events are the most popular form of remote live entertainment in the world, attracting millions of viewers on television, personal computers, and a variety of emerging devices. Sports broadcast are produced using an established and sophisticated process involving producers, directors, commentators, analysts, and a sizeable crew of video and audio technicians using numerous cameras and microphones. This production process continues to evolve as it strives to engage and immerse viewers in the action, suspense, and drama of the remote live event.

In recent years, computer-generated visualizations are increasingly used in sports production to further enhance the viewer experience. Interactive visualization is becoming even more important with the ongoing convergence of television and Internet broadcasting. These trends have led to a significant amount of research activity on sports analysis, visualization, and interactive video browsing [2, 3, 7, 23, 9, 21, 16, 6, 19, 20, 22].

Most work related to sports visualization falls into two categories which significantly enhance the experience of the event:

- Overlay of virtual objects over video using augmented reality techniques – examples of these include the virtual first-down line in football, the virtual offside line in soccer (see [16, 19, 21]), and virtual superposition of video of two competitors taken at different times (see [9, 23]).
- Generation of 3D virtual replays which allow viewing the action from any viewpoint – these include semi-automatic rendering of the action in a virtual environment (see [2, 21]) and reconstruction of a dynamic three-dimensional model of the environment using numerous cameras (as in [8, 17, 10]).

However, the emphasis of these visualization techniques has primarily been on resynthesizing the sport, and not on deeper analysis of the sport. The sports viewer, commentator, analyst, player, or coach is often trying to obtain further insight into performance, style, and ultimately strategy. At the professional level, players are highly trained individuals competing with each other. Subtle changes in performance and strategy can make the difference between winning and losing. Can visualization techniques elucidate and highlight both the typical style and the dynamically evolving strategy of players? Can the type of analysis used in coaching books or playbooks (or more advanced analysis) be automatically derived and displayed during a live sporting event? Can sports action over a period of time be visualized in summary forms that highlight interesting patterns? Is it possible to integrate such analysis and visualization with video browsing to help viewers discover and watch personalized highlights of a sporting event? These are some of the questions addressed in this paper.

Sports can be visualized and analyzed in compelling new ways if sports capture includes the actions of players and the structure of the domain. One way to represent the actions of players is to capture trajectories of motion of the players and the objects influenced by the players. We have developed a system called LucentVision [12] which uses this approach to capturing tennis matches. The system has been used in over 250 international tennis matches since 1998. This paper presents a variety of visualizations made possible by LucentVision using data from some of the tennis matches captured by the system.

2 Overview of the LucentVision System

LucentVision uses eight cameras placed around a tennis stadium to track the players and the ball (see Figure 1). Real-time analysis of video from these cameras determines the motion trajectory for each player and the ball. The details of the video-based tracking appear in [11] for player tracking and in [13] for ball tracking.

To summarize, one camera tracks each player viewing one half of the court. The player tracker segments out foreground/motion regions using differencing operations, tracks local features in the segmented regions, and dynamically clusters the motion of unstable

* This work was done while the author was at Bell Labs.
local features to form a stable motion trajectory corresponding to the centroid of the player.

Six additional cameras track the motion of the ball during serves. The six cameras are divided into four logical pairs and four separate threads track the ball in each serving direction. Ball detection uses segmentation by motion, gray scale, and shape matching. Each thread performs stereo matching for determining 3D position and uses velocity estimation and prediction to constrain the search during tracking. The first thread which succeeds in detecting the ball continues to track the ball in a pair of views and triggers a thread corresponding to an adjacent camera pair when the ball goes out of bounds of its camera pair. A triggering thread passes on the current 3D trajectory to the triggered thread and provides initial expected locations of the ball based on velocity estimation and 3D to image plane mapping. The multi-threaded approach is scalable to many more cameras and to multiple tracked objects.

Techniques from [24] map trajectory data between image coordinates and world coordinates. These motion trajectories are stored in a database along with associated score, environment geometry, location of landmarks, and rules of the game. The broadcast production video is also compressed and stored in the database.

A visualization interface which can reside in different clients, facilitates queries from the database and offers the user a variety of visualizations of the event as discussed in Section 4. The interface is tailored to the computational and bandwidth resources available on different client systems, such as television broadcast systems on site, a remote personal computer with narrowband or broadband Internet connection, or a next generation cell phone.

2.1 Taking Advantage of Domain Structure

The LucentVision system takes advantage of the fact that the dynamics of sports action and the resulting scores in any sport take place within a well-established structure. The sports domain structure includes: a) a well-defined physical environment (court, field, track, arena, etc.); b) well defined landmarks (baseline, service box, net, goal posts, basket, etc.); c) known teams and individuals (often with assigned numbers, uniforms); d) well-defined equipment (rackets, bats, ball, etc.); e) clearly stated rules of action and engagement and definition of legal and illegal actions; and f) specific rules for scoring.

Sports analysis and visualization can be taken to a new level by utilizing this inherent structure. In order to provide interactive insights into sports, we see some key requirements: i) the low level actions by each player need to be identified and derived – what did each player actually do? ii) this action needs to be related to the scores (what were the results of these actions), to the rules and terminology of the game (was that a passing shot, a hand foul, etc.), and to the physical environment (where did this action take place, what was its relation to well-defined landmarks such as a net, baseline, basket, goalpost, etc.); iii) these actions should be time-stamped, and iv) the action data should be organized to support cross-indexing over the different attributes (who, what, where, when, etc.).

Such analysis, performed in real-time, lends great power to users, including broadcasters, fans, coaches and the players themselves. In some sports, even the strategy of the players can be dynamically influenced by such analysis.

2.2 Motion Trajectories for Visualization

As discussed in Section 2.1, a key to providing richer visualizations of sports is to capture the low-level actions performed by the players. These actions include: how they moved their body and their limbs, how they used their equipment (swung a bat, threw a ball, etc.), and how other objects (a ball, a puck, a javelin, etc.) moved as a result of their actions. One way of representing these low-level actions is in the form of motion trajectories, sequences of spatio-temporal coordinates of objects of interest. Although there is some effort towards placing active and passive sensors on the players and objects to facilitate real-time motion capture, most sports do not allow such interference with the players and sports objects. Hence, non-invasive real-time motion capture is needed. Motion tracking of people by camera image sequence processing has been an active area of research in computer vision (see [1] for a review). Real-time people tracking techniques have begun to emerge in the last few years (for example, [18, 15, 5]).
3 Data Selection for Visualization

Once a sporting event is stored in a database in the form of motion trajectories, scores, and other domain specific information, a viewer can explore and interact with the virtual version of the real event. To cope with the sheer volume of captured data, a powerful mechanism of data selection allows the user to choose only the subset of interest. Figure 2 shows the data selection and visualization interface in LucentVision. The interface on the left of the figure is for visualizing player trajectory data and on the right is for visualizing ball trajectory data.

In the context of a tennis match, selection include score-based queries (e.g., all points won by a player against opponent's serve), statistics-based queries (e.g., the points a player ran fastest) or space-based queries (e.g., all points at the net). Each query can be refined further using a time constraint, for example limiting it to one set, one game, or even any particular match period (e.g., the first 30 seconds of the second set). In addition, LucentVision supports historical queries (e.g., all matches between Sampras and Agassi in 1999 that Agassi won). This has become a particularly important feature for tennis viewers, broadcasters, players, and coaches, as LucentVision has already been used to capture data in over 250 international tennis matches.

Each data selection generates an SQL query to the database. Retrieved motion trajectories can be viewed using a number of visualization and animation choices, discussed in the next section.

4 Visualization

4.1 Coverage Maps

The captured motion trajectories for each player can be shown as line drawings mapped onto a three-dimensional model of the tennis court. Figure 3 shows such a mapping for the entire match between Rafter and Kafelnikov during the semifinal of the ATP Championship at Cincinnati in 1999. The virtual court is shown from a top-down view. Although the players change sides on the court during the course of the match, this display shows all the trajectories for one player (gathered from both sides) on one side of the court. Even this simple summary map reveals a good amount of information about how these two players played during this match. Rafter approached the net more often than Kafelnikov. Both players played mostly at the baseline but Rafter was mostly behind the baseline while Kafelnikov was more inside the baseline.

Time spent by a player in different parts of the court (a coverage map) is calculated using:

$$A_{coverage}(x) = \int X_p(t)dt$$  \hspace{1cm} (1)

where $X_p$ represents motion trajectories of the player $p$. In Figure 4, this function is mapped to the virtual court surface for the trajectory data in Figure 3 using a simple color coding. Red indicates higher coverage, followed by yellow, green and blue* This map more clearly highlights the coverage than Figure 3 and reveals additional information. For instance, the color-coded map shows that Rafter's time was more evenly distributed between the forehand (right) and backhand (left) parts of the court while Kafelnikov played more from the backhand side (both players are right-handed).

*In the black and white print, the mapping goes from brightest gray for red to darkest gray for blue.
Figure 5: A coverage map for a subset of a match (points won by Rafter). (See Color Plates for the color version of this figure.)

Figure 6: A coverage map for points won by Rafter in another match. (See Color Plates for the color version of this figure.)

Figure 5 shows a coverage map for a subset of the data shown in Figures 3 and 4, only including trajectories corresponding to the points won by Rafter. Rafter won 61% of the points in the match. Kafelnikov never approached the net during the points he lost while Rafter approached the net a number of times during the points he won. However, the red zones indicate that Rafter still spent more time close to the baseline, and was behind the baseline more than to Kafelnikov. A comparison of Figures 5 and 4 reveals that Rafter spent relatively less time behind the baseline during points he won compared to all the points in the match. This example shows but one of myriad ways in which the data can be broken up and visualized to reveal players’ style and strategy.

Since the LucentVision database archives numerous matches, it is possible to compare a player’s style and strategy in one match with her/his performance in other matches. Figure 6 shows a map for points won by Rafter in an earlier match (the quarterfinal of the ATP Championship at Montreal in 1999) against Kiefer. Comparing Figures 6 and 5, one can see that Rafter approached the net much more in the match against Kiefer in Figure 6 than in the match against Kafelnikov in Figure 5. Also, Rafter was more oriented towards the forehand side of the court in Figure 6 than in Figure 5.

While the preceding examples have shown static coverage maps, these maps can also be animated to show their evolution over the course of a match. Figure 7 shows four frames from one such animation. Each frame also includes the match score.

Figure 7: Evolution of a court coverage map: four frames from an animation.

4.2 End-position Maps

A visualization of the positions of the players at the end of each point, the deciding moment when the point is won or lost, gives another insight into players’ strategies. An end-position map is created by mapping the last node of a player’s trajectory onto the virtual court. In Figure 8, the end-position map is shown for the points won by Agassi against Sampras’ serve in the second set of their encounter during the 1999 World Championships. The two player end positions for each point are linked by yellow lines. Agassi won 33% of the points in the second set for which Sampras served. The figure shows that five out of the eight times that Sampras lost on
serve, he was “caught” at the “T” in the middle of the court. The remaining three times, he was close to the net. The figure also indicates that Agassi mostly hit his winning shots from the backhand side from close to the baseline by hitting “passing shots” as Sampras approached the net. More detailed analysis of players’ strategy is revealed by each pair of corresponding player end positions. Figure 8 also shows the corresponding coverage map. The end position map reveals a significant amount of additional information that cannot be derived from the coverage map.

4.3 Speed Charts

Player motion trajectories provide other information such as the speeds of the players, the distance covered by each player and the directionality of the motion of the player. Figure 9 shows a graph of the speed of each player over the course of the match. The speed is calculated as:

$$A_{\text{speed}}(t) = \frac{dX_p(t)}{dt}.$$  

where $X_p$ represents motion trajectories of the player $p$. The speed chart in Figure 9 shows the moving average of the player speed over every two second period during play.

Such graphs give insight into player stamina and fitness. In Figure 9 it is seen that Agassi is faster than Sampras in the first set, but slows down significantly in the second and especially, the third set. In contrast, Sampras speeds up significantly in the third and deciding set.

4.4 Virtual Replays of Serves

A tennis ball may travel at speeds as high as 235 km/h (140 mph) during serves, which makes it hard to follow on video alone. As performance on serves often decides the performance over the entire tennis match, a visualization derived from the three-dimensional motion trajectory offers valuable additional insights. Figure 10 shows a virtual replay of an ace served by Venus Williams during the 2000 US Open final. The virtual environment is constructed from a three-dimensional geometric model, texture mapped with images of the court and the audience, using VRML, OpenInventor and MasterSuite tools. The virtual replay environment allows a replay of a serve at any speed from any viewpoint with or without a display of the ball’s trajectory. The viewpoint shown in Figure 10, which represents a position in front of the receiver, cannot be obtained from any camera on site. The figure also shows some statistics - the ball speed at the racket, after the bounce and the height at the receiver, all calculated from the three-dimensional ball trajectory data. The receiver’s position is known from the player tracking system. Virtual replays show ball placement while bounce and speed statistics give an idea of the style of the serve (top spin, etc.).

Another visualization revealing the style of play is a sequence of virtual replays of serves for any portion of a match. Figure 11 shows replays of all aces served by Safin during the final of the
4.5 Service landing positions

Service placement offers another insight into a player’s strategy. Analysis of the 3D ball trajectory, with appropriate interpolation, yields the ball landing position for each serve. If the 3D trajectory of length \( n \) has time samples \( (t_1, t_2, \ldots, t_n) \), and the time sample \( t_e \) represents the last sample with a negative \( z \) velocity (computed from time \( t_{e-1} \) to \( t_e \)), then the landing position is at a time \( t_f \) which is either between \( t_e \) and \( t_{e+1} \) or between \( t_{e-1} \) and \( t_e \). In the first case, forward projection from the 3D velocity and acceleration parameters at time \( t_e \) determine when the ball reaches the ground. In the second case, backward projection from the velocity and acceleration parameters at time \( t_{e+1} \) determine the landing location and time. We choose one depending on how well the velocity at the interpolated position matches the velocity at the tracked positions.

Figure 12 shows service landing positions for both players (Venus Williams and Martina Hingis) for the semifinal of the US Open 2000. First serves are shown in yellow (light) while second serves are shown in blue (dark). This is another example of a single picture which reveals a significant amount of information about the match. Hingis has many more first serves than Williams. However, her first serves are conservative – landing closer to the center of the service box, rather than at the corners. Williams’ first serves are more targeted towards the corners of the service box. Williams’ second serves are much more conservative. The total number of serves for Hingis are greater than Williams, indicating that Hingis lost more of her service points and had to serve more.

4.6 Combined Use of Multiple Visualizations

Significant insights into performance and strategy is derived by using multiple visualizations from one or more matches. The power of the LucentVision system is that it gives the user the ability to explore and integrate different visualizations to find interesting information. Figures 13 and 14 illustrate this. The top portion of Figure 13 shows the coverage map for points won by Sampras against Agassi’s serve in a match in the second round of the World Championships in 1999. The figure indicates that Sampras wins only 23% of the points against Agassi’s serve. The coverage map shows

Sampras almost entirely on the backhand side and close to the baseline. The lower part of Figure 13 shows the service landing positions for the serves by Agassi. Agassi has a high percentage of first serves and remarkably targets his serves on the corners of the service box, particularly on the backhand side. There is only one serve on the forehand side in each service box! This explains the coverage of Sampras and suggests that this high-precision serving lead to Agassi’s convincing victory.

This can be contrasted with Figure 14, showing similar maps when the same two players met in the final of the same tournament. Agassi’s serving is not as consistent as in the previous match, both in placement and in the percentage of first serves. This offers Sampras the opportunity to move more freely on the court resulting in the defeat of Agassi.

4.7 Integrated Visualization and Video Browsing

The visualizations discussed in the preceding sections can also be enhanced with video browsing for retrieving interesting video clips. Video indexing and retrieval is an active area of research [4, 26, 25, 20]. This research work is primarily focused on segmenting and classifying broadcast video. Attaching domain semantics to the broadcast video is the biggest challenge for these approaches. LucentVision tightly integrates domain structure with the capture of video and motion data. Hence, the system can serve as a very powerful video retrieval system.

Figure 15 shows an example of combining visualization with
video navigation. In the top portion of the figure, we find a visualization of landing positions of all aces served by Sampras during the match. The user selects the region corresponding to the left corner of the left service box. This results in the selection of the three landing positions shown in the left picture in the second row of Figure 15. Frames from each of the corresponding three video clips appear in the second and third rows of Figure 15. The video clips also show the corresponding game score. The three selected landing positions (top to bottom) correspond to a point in the 4th game of the second set, 5th game in the first set, and 7th game in the first set, respectively. Immediate non-linear access to these three different video clips is made possible by this integrated visualization and video browsing. For more detail on this aspect of LucentVision, see [14].

5 Conclusions

This paper presents a novel visualization scheme for sports based on motion trajectory data for the players and other objects such as a ball. The LucentVision system for tennis has illustrated some of the numerous visualizations made possible by this scheme. The examples show that "a picture is worth a thousand words" - a single summary visualization can reveal a remarkable amount of information about the performance, style, and strategy of the players.

Future research needs to address motion tracking and visualization in more complex sports involving multi-player coordination and interaction. Another significant research problem is to track the motion of individual limbs of players to follow their actions in greater detail. New visualizations need to be developed to illustrate the subtleties and changes in a player’s style. The primary challenge in sports visualization is showing what is important and interesting in a form that is very easy to understand.

The LucentVision system has been used extensively in international tennis matches (more than 250 matches). Television commentators have found that the visualizations provided by LucentVision illustrate what they are saying and help them better explain the sport. Many coaches and players who have seen the system also recognized its value in analyzing their matches.

References


Figure 15: Integration of visualization and video browsing. (See Color Plates for the color version of this figure.)