Game, Shot and Match: Event-based Indexing of Tennis

Damien Connaghan, Philip Kelly and Noel E. O'Connor
CLARITY: Centre for Sensor Web Technologies
Dublin City University
Dublin 9, Ireland
connagha@eeng.dcu.ie

Abstract

Identifying events in sports video offers great potential for advancing visual sports coaching applications. In this paper, we present our results for detecting key events in a tennis match. Our overall goal is to automatically index a complete tennis match into all the main tennis events, so that a match can be recorded using affordable visual sensing equipment and then be automatically indexed into key events for retrieval and editing. The tennis events detected in this paper are a tennis game, a change of end and a tennis serve - all of which share temporal commonalities. There are of course other events in tennis which we aim to index in our overall indexing system, but this paper focuses solely on the aforementioned tennis events. This paper proposes a novel approach to detect key events in an instrumented tennis environment by analysing a players location and the visual features of a player.

1 Introduction

In this paper, we describe our approach to automatically index a tennis match into a hierarchy of events. Each of these events is defined by the official rules of tennis [1], which are governed by the International Tennis Federation (ITF). Rather than indexing a televised broadcast of a tennis match, we are detecting events from an indoor tennis match, where a number of cameras are positioned around the court. Multiple viewing angles provide a number of viewing perspectives on a match and the cameras used are also suitably priced for any local tennis club who may wish to enhance their coaching facilities on a cost effective budget.

The tennis events which we wish to automatically index for our visual coaching system are a match, set, game, change of end, rally and type of tennis stroke (serve, forehand and backhand). However, in this paper, we detect the following events only: serve game and change of end. Automatic detection of all these key events is necessary to reduce the time a coach will spend manually indexing a recorded match. Event detection also goes one step further to achieving our goal of automatic video production of a tennis match. By this, we mean the ability of our video coaching system to decide which available camera(s) offer the best viewing perspective to capture a given tennis event.

In previous publications, we outlined our work developing a suitable interface, where a coach can quickly retrieve key events from a tennis match, however this paper describes the algorithms that are used to automatically index these key events [3].

2 Related work

There has been many advancements of late in event detection within sports video, many of which are in broadcast video for field sports, such as soccer and American Football. One such work, which was published by Sadlier el. al. [11], developed an audiovisual feature based framework for event detection in broadcast video of multiple field sports. Features for high level events are selected and detectors are built for each event. These features are rooted in characteristics common to all genres of field sports. In [11], the evidence gathered by the feature detectors is combined by means of a support vector machine, which infers the occurrence of an event based on a model generated during a training phase. The system was tested generically across multiple genres of field sports including soccer, rugby, hockey, and Gaelic football and the results conclude that high accuracy event retrieval is achievable.

Tennis however, has not achieved the same volume of research within event detection as soccer or American football. In fact, quite a lot of published research in this area which relates to tennis, describes approaches for detecting tennis strokes from broadcast video only. Bloom and Bradley [2] detected a shot *keyframe* when the ball collides with the racket and determined stroke classification from heuristics based on the player and racket locations on impact. In [13], a tennis match was recorded in broadcast video and player motion was detected by extracting optical flow features. Support Vector machines (SVMs) were then used to classify the tennis strokes into either a forehand or backhand using a left-swing or a right swing class.

Petkovic et al. [10] uses *pie* features and several Hidden Markov Models to classify tennis strokes into either forehand, backhand, service, smash, forehand volley and backhand volley. Shah et al. [12] model a tennis player's body as a skeletonisation and feed an orientation histogram of this skeleton into SVN classifiers to distinguish forehand, backhand and 'neither'.

In our previous work to detect tennis strokes in an instrumented environment [8], we used cameras positioned behind the player whereby for each frame in a tennis shot, the player is segmented into a foreground region and then sliced into pie segments to extract action features. These features were then classified into serves, backhands or forehands using either SVM classifiers or K-means nearest neighbour clustering.

The final system we are developing will detect all the key tennis events, such as a game, set, change of end, rally and various tennis strokes from an instrumented environment. The volume of published work on detecting key tennis events other than tennis strokes within an instrumented or broadcasted environment is quite sparse.

3 Infrastructure Overview

As described in detail in [4], we have instrumented an indoor tennis court with nine IP cameras positioned around the court, with pan, tilt and zoom (PTZ) capability (see figure 1 for the location of the cameras around the court and a sample of the different camera views available). The two cameras at the centre of the baseline at either end of the court are AXIS 215 PTZ cameras, which are positioned 2.8 meters above the court and have very high zoom functionality, as well as physical pan and tilt. The high zoom is useful for obtaining a front view of the opposing side, or for focussing on the feet of the player from behind the court baseline. The other seven cameras are AXIS 212 PTZ cameras which have wide angle lenses (140°) and include fast digital PTZ functionality by subsampling from a high-resolution sensor. Six of these cameras are located at 3.5 meters above the court, while one overhead camera is positioned at 13.8 meters from the ground. All the cameras have a resolution of 640×480 and a frame rate of 30Hz.

4 Tennis Events

A tennis match consists of a series of high-level events which can be structured in a hierarchical manner as shown in Figure 2(b). To help segment a tennis match into key

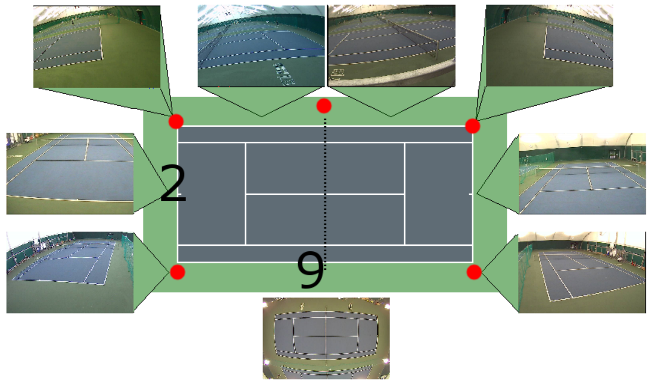


Figure 1. Camera locations, cameras 2 and 9 are used for event detection.

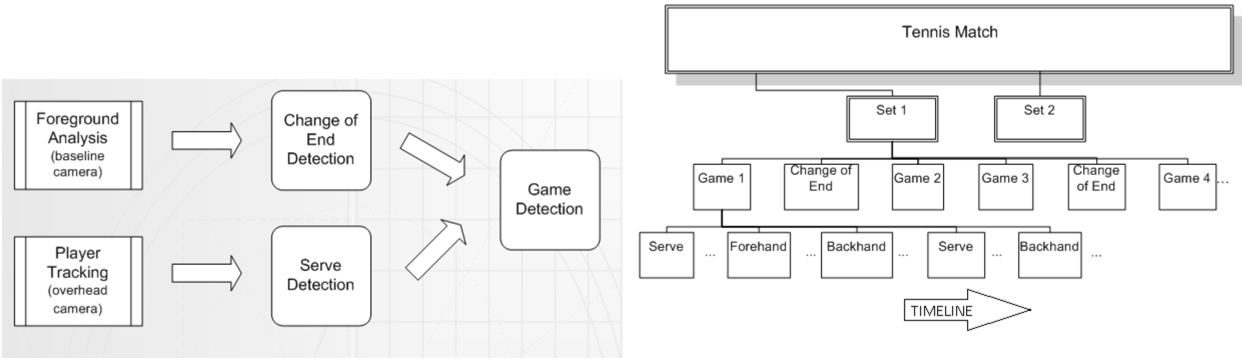
events we have defined a framework, which defines all the key events of a tennis match, such as a set, game, stroke, rally etc. Each of the events in this framework are aligned with the official rules of tennis [1].

- **Stroke** is the act of hitting a ball with a racket, which at a sub level is either a **forehand**, **backhand** or **serve**.
- **Rally** is continuous period of play starting with a serve and ending when one player scores.
- **Game** is a series of rallies and strokes, culminating in one player reaching a score of 4 and winning by 2 clear points.
- **Change of end** occurs after the first game in the first set and then after every two games thereafter.
- **Set** is a series of games, ending when one player wins six games and leads their opponent by two games or more , or when a player wins by two clear points in a tie breaker.
- **Match** is a series of sets, ending when one player wins a given number of sets.

In this paper, we present our research findings for detecting the following events: games, change of end and serves.

5 System Overview

The visual coaching system being developed by us will ultimately index all the tennis events shown in Figure 2(b). In this work, however we focus on detecting three tennis events, a serve, a change of end and a game, as shown in the processing chain in Figure 2(a). The following sections discuss the processes used to index each of these three tennis events.



(a) (b)

Figure 2. (a) System Processing Chain: Processes which detect tennis events have double vertical lines. Serve, change of end and a game are the tennis events detected. (b) Tennis events flow in a top down hierarchy.

5.1 Player Tracking

In [9], we published our results on the development of algorithms for extracting useful metadata from the overhead court camera (camera 9 in Figure 1). Using background subtraction and hysteresis-type blob tracking, we can track the tennis players positions around the court. The performance of the module was evaluated using ground-truthed data and the results were very promising, providing a wealth of data on both of the players locations during a match. The player tracking module in [9] is used in this paper to develop a new approach for serve detection as described in 5.4 .

5.2 Player Foreground Extraction

To detect if the players have switched sides we first extract the player as foreground from camera 2, which is positioned behind the baseline, as shown in Figure 1.

Since we aim to extract features from the player foreground, shadows need to be removed and for this reason we use a layered background model that includes a module for shadow removal [8]. Shadow pixels are calculated for each frame by detecting if the difference between foreground candidate pixels in the current frame, f_i , and the corresponding pixels in the background model, b_i , is below a user defined threshold t. A pixel, (x, y), is declared as shadow if

$$|fi(x, y) - bi(x, y)| <= t$$

otherwise the pixel declared as a foreground pixel.

An example of two player foregrounds can be seen in Figure 4. Once a player foreground has been extracted, we can then detect a change of end as described in the next section.

5.3 Change of End Detection

The rules of tennis as laid down by the ITF [1] state that "players shall change ends at the end of the first, third and every subsequent odd game of each set."

To detect a change of end event, we sample camera 2 (see Figure 1) at i intervals. We then compare the player features for the image located at position y, to each of the previous n sampled images, to detect if large changes have occurred in the player's appearance. Player features are extracted from each image sample by firstly extracting the player as foreground and then splitting the resulting HSV image into three channels (Hue, Saturation and Value). We discard the first two channels (Hue and Saturation), as analysis concluded that the Value channel alone provides sufficient information to accurately detect a change in the players appearance.

We then create an image histogram from the Value channel, which represents the brightness in the player foreground. The similarity distance between the histogram at position y and each of the previous n sampled histograms is calculated using the Bhattacharyya coefficient, which has been widely used to compare image histograms [7]. Where the similarity distance between two histograms exceeds a given threshold T for x consecutive samples, a new player is deemed to be present in the image at position y and therefore a change of end event has occurred. To detect if a different player is in camera shot, this approach is greatly aided by a clear player foreground as shown in Figure 4.

The following parameters were used to detect a change of end event. An image from camera two is sampled at 2 second intervals. If an end change can be detected to

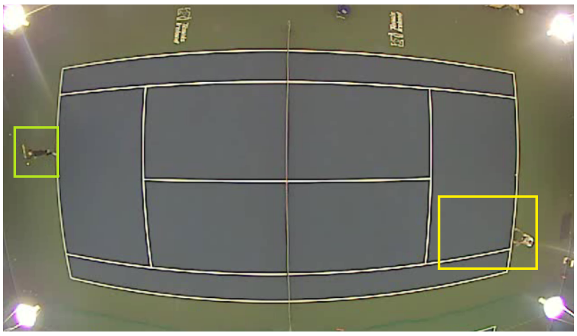
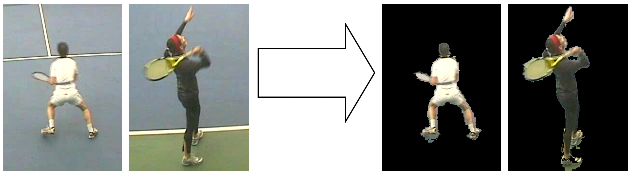


Figure 3. Player A is inside serve zone (left side) for 2 seconds and Player B is inside return zone for four seconds, therefore Player A is serving.



(a) (b) (c) (d)

Figure 4. Player foreground is extracted and the colour features of the players are compared using the Bhattacharyya coefficient to detect a change of end event.

within 2 seconds this is sufficient for our system. Where the Value histogram of the current sample has a similarity distance greater than .33 to each of the previous 3 sampled histograms, it is declared that a new player is in the current frame at position y and therefore this temporal position is flagged as a change of end event. The similarity distance of .33 was chosen by analysing the similarity distances for contrasting players from various tennis matches. Analysis of different tennis players concluded that slight differences in a players appearance can be detected. Since only two players are involved in a match, this change of end detector works very well in this environment.

5.4 Serve Detection

Using the data from the player tracks in Section 5.1, we can locate both players positions and map these coordinates to the tennis court to determine each player's location on the court. Then, by determining both players locations on the court at all times during a match, we are able to recogonise a serve event.

Conroy et. al. [6] states that a "receiving player is usually beyond the baseline on the opposite side of the court. However in practice, the receiver can be closer to the net, or within the baseline". A server though, is always located behind the center of the baseline and will alternate serves between the left and right of the center mark on the baseline. Taking these starting positions for a server and receiver into account, we created two distinct zones to detect a serve event, one zone to locate a server (serve zone) and the second zone for the receiving player who returns the ball following a serve event (return zone). By observing video footage of where tennis players are located on court whilst a serve event is in progress, we conclude that a serve will originate from four unique serve zones and a serve will be returned by the opposing player from four unique return zones. There are eight zones used to identify serves, four are serve zones (S_1, S_2, S_3, S_4), whilst the remaining four zones are return zones (R_1, R_2, R_3, R_4), as illustrated in Figure 5.

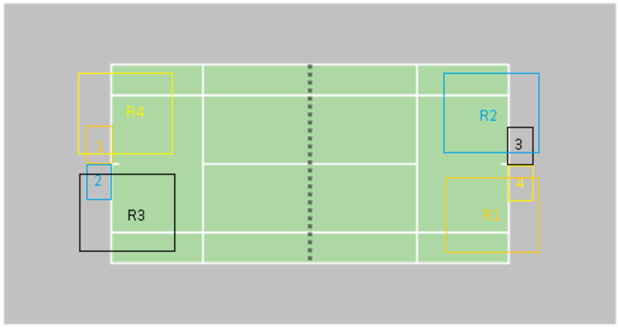


Figure 5. Serve zones 1, 2, 3, 4 and the corresponding return zones R_1, R_2, R_3, R_4 are used to detect a serve based on both player's locations.

The rule for serve detection is as follows. For two players P_1 and P_2, let P_1^t and P_2^t be the positions of P_1 and P_2 at time t respectively. If P_1^t is within the bounds of the serve zone S_x for s seconds and P_2^t is within the bounds of the return zone R_x for q seconds, then P_1 is declared as serving from S_x. A similar rule exists to detect if P_2 is serving.

Serve zone coordinates were obtained by observing the typical locations of players when serving from each serve zone. A similar approach was used to detect return zone coordinates. To obtain efficient coordinates, a number of evaluations were carried out whereby, we manually tested an initial set of coordinates based on observations and by analysing the precision and recall results, we modified the respective zone coordinates until the error in the precision and recall results were minimised. The evaluation results for this algorithm are given in Section 7.

5.5 Game Detection

In this section, we provide a solution for segmenting a tennis match into units of tennis games. From the official rules of tennis [1], a standard game of tennis is described as an event which "contains a count of the total number of points won by each player. When a player's score reaches the fourth point and they are winning by two clear points, the game is won by that player. If the scores don't have two clear points between them when the fourth point is reached, the game continues until one player takes the lead by two clear points."

Automatic indexing of the game event, however, requires careful analysis of the key characteristics which generally occur during this event. One such characteristic of a game, is that only one player may perform a service during a game. The serve itself alternates between players after each game, so identifying the origin of a serve, using the serve detection algorithm in Section 5.4 will provide the first cue needed for game detection.

A second cue which helps identify a game is the change of ends rule. This rule forces the players to change ends after the end of the first game and then after every two games thereafter [1]. This means that for the first two games, all serves will originate from the same end of the court. This is because one player will serve the first game, then a change of end event occurs and the second player will serve the second game from the same side as the first player.

Taking these two characteristics into account, we have developed a novel approach which combines the results from the serve event (Section 5.4) and the change of ends event (Section 5.3) to detect the start and end of all games. By detecting which end a serve originates from and combining this data with which player is located inside the live serve zone, we can detect the beginning of a game boundary at time t. The end of the previous game boundary is detected by finding the last serve which occurs before t as described in Section 5.4. The classification results for game detection are analysed in Section 7.

6 Data Capture

In previous capture sessions we captured data from advanced tennis players and classification results of tennis stroke recognition was published in [8] [5]. As the high level of classification proves, extracting features from advanced tennis players and recognising the tennis strokes played can be achieved by inferring features from the video.

In this data capture however, the detectors are applied to two tennis matches, which were played by different classes of players. Match 1 contains two sets of tennis, from two advanced tennis players. Match 2 also contains two sets of tennis but is contested by two amateur tennis players.

Since Match 1 was played by regular players, the match was allowed to continue without the intervention of a referee and coincidentally the players did not perform a change of end event at all the required times. This decision was made by the players in order to speed up play and it is quite common to do so in a non competitive tennis match. However, as the results prove, our indexing system can still detect a new game in this eventuality. To achieve continuity in the tennis events of Match 2, a third-party referee was present to advise the amateur players where necessary.

There are new challenges to be explored in finding patterns in the kinematic movements and locations of amateur tennis players. This is because an amateur player will not exhibit the same patterns of movement between consecutive strokes, whereas an advanced tennis player will retain consecutive movements when performing the same tennis stroke.

The duration of Match 1 is 39 minutes and contains 376 ball hits, 80 serves, 14 games and 3 change of end events. The duration of Match 2 is 51 minutes and it contains 557 ball hits, 142 serves, 15 games and 7 change of end events. In Match 1, Player 1 wore a yellow shirt with dark shorts, while Player 2 wore a grey top with dark shorts. In Match 2, Player 1 wore white shorts with a white top and Player 2 wore dark top with dark shorts.

7 Event Detection Results

The event detectors were trained on Match 1 and when high precision and recall results were obtained, the event detectors were then performed on Match 2 without adjusting any parameters. To create a ground truth, every event in the complete dataset, which was to be automatically indexed by the system, was manually annotated in both matches.

Event	No. in dataset	Precision	Recall
Match 1 (Training data)			
Games	14	1.0	1.0
Change of Ends	3	1.0	1.0
Serves	80	.80	.89
Match 2 (Test Data)			
Games	15	1.0	1.0
Change of Ends	7	1.0	1.0
Serves	142	.82	.86

Table 1. Precision and recall results for both tennis matches, which were contested by different players.

Although the serve detectors performs very well, we hope to improve on the serve detection accuracy. Analysis of the errors made by the serve event detector, point to there being a noisy player tracking data. Sometimes the player track terminates when the player is outside the field of view or is stationary for a short time. Improvements in player tracking will help to improve serve detection.

The time when both players change ends of the court is required to detect which player is serving, which is a vital cue for detecting game boundaries. This provides us with a simple player identification and can be used to detect what player is serving and what player is returning at any time in the match. If a new player was detected in camera 2 and this event was correct to $\pm$ 2 seconds, the result was recorded as correct. As the results from both matches illustrate, a change of end event can be detected to 100% accuracy.

In Match 1, the players did not change ends at the required times. In this event, the origin of the serve will change from one side of the court to the other, from game to game. Therefore, to compensate for a non-change of end event, we simply use the serve origin as a cue to detect a new game boundary.

As the results illustrate, game boundaries are successfully detected. The high accuracy in game boundary detection is directly related to the high accuracy also achieved in detecting a serve event and change of end event, as shown in Table 1. If the system detected a game boundary between the final serve a previous game and the first serve of a new game, the game detection was marked as correct.

8 Conclusions and Future Work

The results in the previous section are very encouraging. Although the change of end detector performs very well, future work will extract additional features from the player foreground to detect a new player, such as player height while located at the baseline or analyse the colour regions within the player foreground. Another solution to detecting a change of end event could be inferred from the overhead camera. Though currently, individual player tracks tend to overlap during a change of end event. This is because both players may walk past each other and the tracks may overlap.

Quantitative evaluations have concluded that this event indexing system performs well, even with standard video equipment. The next step in the development of this indexing system is to merge our stroke recognition detectors with these event detectors, which will bring us towards a complete video indexing system for tennis coaches.

9 Acknowledgments

This work is supported by Science Foundation Ireland under grant 07/CE/I1147.

References

[1] International tennis federation(itf) - rules of tennis. http://www.itftennis.com/abouttheitf/rulesregs/rules.asp.

[2] T. Bloom and P. Bradley. Player tracking and stroke recognition in tennis video. In *Proceedings of the WDIC*, pages 93–97, 2003.

[3] D. Conaghan and N. E. O'Connor. An intuitive user interface for visual sports coaching. In *Proceedings of the Irish Human Computer Interaction (iHCI) Conference*, pages 141–142, 2010.

[4] D. Connaghan, S. Hughes, G. May, K. O'Brien, P. Kelly, C. Ó Conaire, N. E. O'Connor, D. O'Gorman, G. Warrington, A. F. Smeaton, and N. Moyna. A sensing platform for physiological and contextual feedback to tennis athletes. In *Body Sensor Networks (BSN) Workshop*, pages 224–229, 2009.

[5] D. Connaghan, C. O'Conaire, P. Kelly, and N. E. O'Connor. Recognition of tennis strokes using key postures. In *ISSC 2010 - 21st IET Irish Signals and Systems Conference*, Cork, Ireland, 23-24 June 2010 2010. Institution of Engineering and Technology, Institution of Engineering and Technology.

[6] K. Conroy and M. Roantree. Enrichment of raw sensor data to enable high-level queries. In *Proceedings of the 21st international conference on Database and expert systems applications: Part II*, DEXA'10, pages 462–469, Berlin, Heidelberg, 2010. Springer-Verlag.

[7] K. Ke, T. Zhao, and O. Li. Bhattacharyya distance for blind image steganalysis. *Multimedia Information Networking and Security, International Conference on*, 0:658–661, 2010.

[8] C. Ó Conaire, D. Connaghan, P. Kelly, N. E. O'Connor, M. Gaffney, and J. Buckley. Combining inertial and visual sensing for human action recognition in tennis. In *Proceedings of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams*, pages 51–56, 2010.

[9] C. Ó Conaire, P. Kelly, D. Connaghan, and N. E. O'Connor. Tennissense: A platform for extracting semantic information from multi-camera tennis data. In *International Conference on Digital Signal Processing (DSP)*, pages 1062–1067, 2009.

[10] M. Petkovic, W. Jonker, and Z. Zivkovic. Recognizing strokes in tennis videos using hidden markov models. In *VIIP*, pages 512–516, 2001.

[11] D. Sadlier and N. O'Connor. Event detection in field sports video using audio-visual features and a support vector machine. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(10):1225–1233, 2005.

[12] H. Shah, P. Chokalingam, B. Paluri, S. N. Pradeep, and R. Balasubramanian. Automated stroke classification in tennis. In *ICIAR*, pages 1128–1137, 2007.

[13] G. Zhu, C. Xu, Q. Huang, W. Gao, and L. Xing. Player action recognition in broadcast tennis video with applications to semantic analysis of sports game. In *MULTIMEDIA '06: Proceedings of the 14th annual ACM international conference on Multimedia*, pages 431–440, New York, NY, USA, 2006. ACM.