CSC 449: Machine Vision (Final Project Report)
Court Positioning Analysis in Tennis
Team: Vision-aries

Sayudh Roy
Department of Computer Science
University of Rochester, NY
sroy8@ur.rochester.edu

Deepak Pandita
Department of Computer Science
University of Rochester, NY
dpandit2@ur.rochester.edu

Abstract

In this project, we develop a real-time player tracking technology for broadcasting applications in the sport of tennis. The system is able to track and record the locations and motion of tennis players whilst the live point (or a rally) is in progress and be able to in turn, provide a number of statistics about the game, for the benefit of the performance of the players. From these trajectories, the essential information extracted are the court coverage patterns, the distance travelled by the player and the instantaneous speed and acceleration at any time instant. Although with recent developments in tracking people in real-time is not too novel, what was challenging in the project was to be able to detect the exact court from a broadcast and then be able to map it to their two-dimensional counterparts for statistics calculations.

1. Introduction

In tennis, for a player to be able to position himself before playing his next shot is of utmost importance. With the ability to position himself/herself on court correctly while anticipating where the opponent is going to play the ball next, can be crucial to his performance. In preparation for a match, if a player can review what the general trend or pattern his/her opponent follows while returning, it will help decide where to place himself/herself after playing his/her current shot and before the next one. The objective hence, is to recover with more intensity and adjusting the position based on the location of the ball being played.

With the help of computer vision techniques, data analytics for a fast-paced sport like tennis, is useful and can be utilized for strategic development of the players. With most technologies focussing on ball trajectories and statistics on how the ball behaves throughout the game, we in turn have decided to concentrate on finding out more about the players and their movements throughout. This will not only help, a certain player perform well by looking at his own coverage but also give an insight as to his opponent's style of play.

2. Related Work

Person detection and tracking analysis has been receiving increasing attention from computer vision researchers. With a wide plethora of applications, such as surveillance, athletic performance etc., this interest is being motivated even further. In terms of the detection and tracking algorithm we use the person detector algorithm from Dalat et al.[1] to detect the players in the images. The method is based on evaluating the locally normalized histograms of image gradient orientations in a dense grid. These normalized descriptor blocks are called Histogram of Oriented Gradient (HOG) descriptors. These are used as features for a Support Vector Machine (SVM) classifier which is trained for pedestrian detection in city scenes. This method performs consistently better than the wavelet based and Scale Invariant Feature Transform (SIFT) descriptors.

3. Problem

Given a video of a tennis rally (or game/match), we track the position of the players and generate a heat map based on their exact locations on (or beyond the baseline of) the court. This helps to analyse certain patterns of play like court coverage patterns, of different players. These include where the player usually positions while serving (which in turn, may determine where to serve; near the "T" or wide). Also, it helps to understand if a player tends to spend more time close to the net or behind near the baseline. Strategic development to face a particular opponent is crucial, and our project aids in that respect. We will also analyse the patterns of change in speed at which the players move as the rally (or game/match) progresses. Certain players may slow down after a series of shots, which can be used to ones advantage on further analysis.

4. Method

We use a standard approach to detect and track the movement of the players. To generate the heat map on the court, we detect the corners of the court and do a projective transformation to identify the exact location of the players relative to the court. Further, on tracking the players movement over frames in the video, we are able to calculate the distance he/she travels during the course of a point or rally. On calculating the first derivative, we are able to generate the player's velocity and the second derivative gives the acceleration.

Therefore, to summarize, we have the following steps in our project:

- Detect the tennis players
- Detect the court edges (or lines)
- From the edges (or lines), detect the corners
- Extracting the corners of interest only
- Creating 2D map of court and mapping the actual court to the map and mapping the detected tennis player onto that space
- Computing the statistics of the player with heatmap depicting court coverage

The project code has been developed on MATLAB with its Image Processing Toolbox.

5. Experiments

The very first task for this project is to able to detect the tennis players from a live broadcast video. For this purpose, we use a people detection algorithm which is trained on an Support Vector Machine which is able to detect unoccluded people in an upright position using the Histogram of Oriented Gradient (HOG) features. Since the players to be detected in the video frames are quite small, the model which is trained on small image sizes (96 by 48 pixels) was used. The output achieved after applying the raw model on the data is shown in figure 1.

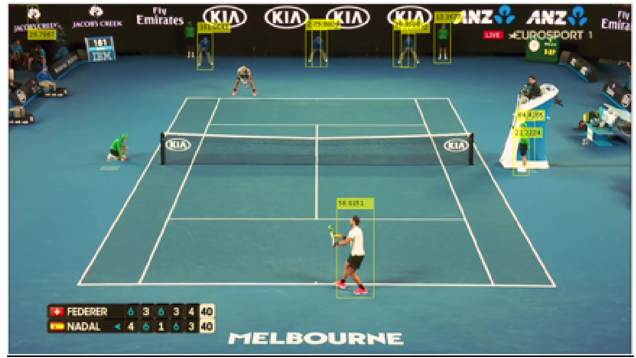

Figure 1. Initial Player Detection

One of the major limitations in this step, which is clearly visible in the image above iss the fact that the algorithm is able to detect a number of other people in the image too, namely the chair umpire, the lines umpires and the ball boys. To overcome this problem, a number of different options are explored. We decide to threshold the detector higher so that it is able to detect people with higher confidence values. But the trade-off for this is that oftentimes, when the player to be tracked is not in an absolute upright position, the detector fails to capture his location. But this is countered as we record his positions over frames and hence, even if the detector fails to capture a particular frame, over time his locations are stored. Also, since for this project we are focussing on only the player on the closer end of the broadcast, we decide to clip the video frames (using the court interest points, which are discussed later), to create a region of interest, only from where the player is to be detected. This results in the output shown in figure 2 on one of the frames.

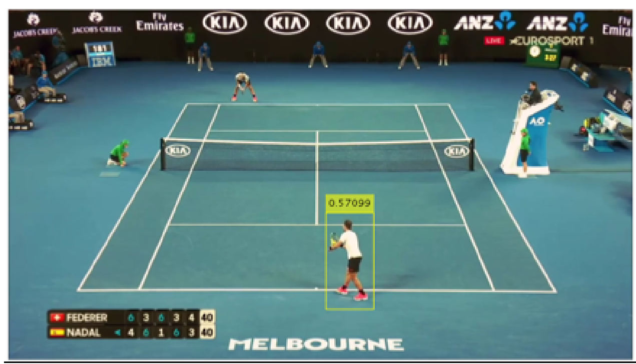

Figure 2. Player Detected

The next few steps comprise detecting the court lines (or edges) and being able to find the corner points of interest only. A slight advantage we have about the sport of tennis is the fact that during a broadcast, the camera is almost always able to capture the entire tennis court and is usually placed on the centre axis of the court.

We traverse over the frames of the image and for each frame it is first converted to grayscale for feature detection. The Canny edge detection algorithm is then used to obtain the discrete edge points from the image frame. Arguably, we could say that the line equations could be extracted from such edge points itself, but since the broadcast input is not vectorized, its raster forms do not allow line feature extractions as the edges are jagged due to the image pixels. This can be seen in figure 3.

As explained above, to detect the corners of the court,

Figure 3. Canny Edge Detection

only using a conventional edge (Canny) or corner (Harris) detector is not sufficient due to the rasterized images. Even though it is able to detect certain lines or corners, it is almost impossible to be able to differentiate it from the noisy data it produces (from the advertisement boards and other environmental instances). To overcome this limitation, we exploit the use of Hough transforms (a quite fast and accurate algorithm), which is a parametric representation of a line in two-dimensional space. It uses the discrete edge points detected by the Canny edge detector to merge them into lines and generalize the straight-line features. It adopts a parametric line equation to identify all possible orientations defined by (theta, rho) at a given point (x, y). With Hough transforms we are able to detect all possible lines which are present in a given video frame. Appropriate thresholding is applied to able to detect only the eight lines which are needed for detecting the points of interest. It can be viewed as the representation in figure 4.

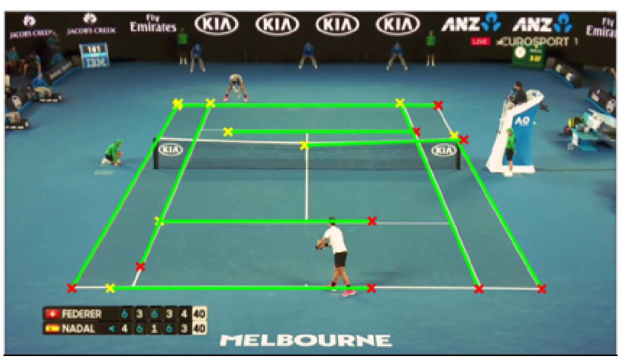

Figure 4. All Lines

With the eight lines of interest stored per frame, the next task is to be able to detect and extract just the four corners of the court which is then applied to the projective transformation. To achieve this, a function is responsible to accept the equations of any two lines (which are computed using the resultant Hough features from the previous step), and solve them to find the corresponding corner points. The sixteen points extracted are as follows, after which to get the four corners is just by thresholding the x and y coordinates of the points. This is shown in figure 5.

Figure 5. Court Corners

Our step is to perform projective transformation with the extracted corner points to convert it into two-dimensional space with actual tennis court dimensions. A standard tennis court is given in the figure 6 (Credit: allcourtdimensions.com).

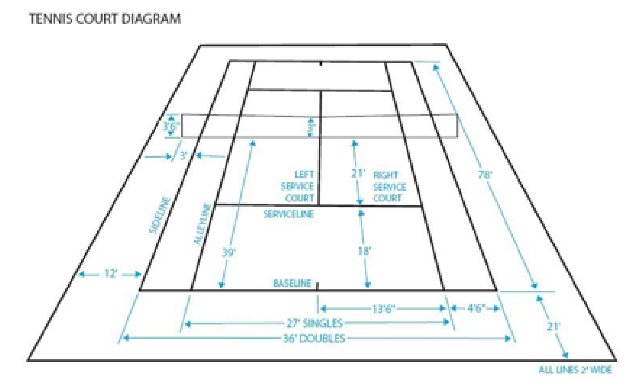

Figure 6. Tennis Court Dimensions

Therefore, a regular tennis court is of the dimensions, 78 by 36 feet. According to our model, 78 feet is represented by 390 pixels, while 36 feet by 180 pixels. Using these values for the transformation, the output shown in figure **??** is generated after applying geometric transforms on the video frame(s). With the calculated location of the player at the given video frame, the coordinates of the player are then passed on to the transformed matrix and forward geometric transformation is applied to estimate the players location with respect to the transformed image. Since, the player closer to the camera is being tracked and in tennis most often, players tend to remain outside the baseline, a court offset of 80 pixels (or 16 feet) is applied.

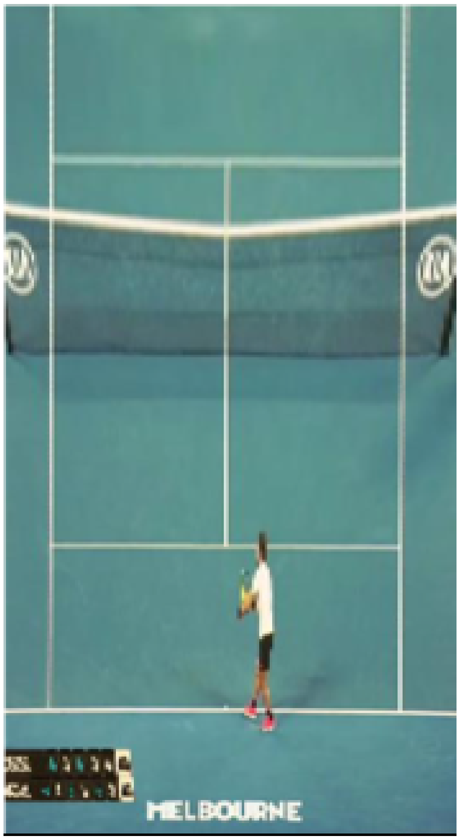

Figure 7. Transformed Image

After performing the aforementioned steps, the last step is to calculate the statistics over a single point. For the purpose of this demonstration, we compute the statistics of Rafael Nadal during a rally (one of the longest and most crucial in the match) in his final against Roger Federer in the Australian Open 2017 final.

From the detected player locations, we are able to calculate the distance travelled by the player by computing Euclidean distance between every subsequent point and applying appropriate scaling to convert pixel distances to feet. Since for a video, we know its frame rate, it is possible to be able to calculate the time elapsed at a certain instant. Instantaneous speed and acceleration are calculated by computing first and second derivatives of the distance over time. For example, at a time instance, t = 7.51 seconds, the calculated statistics are shown in figure 8.

After the course of the entire rally the following are the final statistics calculated:

```
Approximate Distance Covered:  821.05 FT
Approximate Time Elapsed:  35.50 SECS
Approximate Speed:  25.38 KMPH
Approximate Maximum Speed:  27.95 KMPH
```

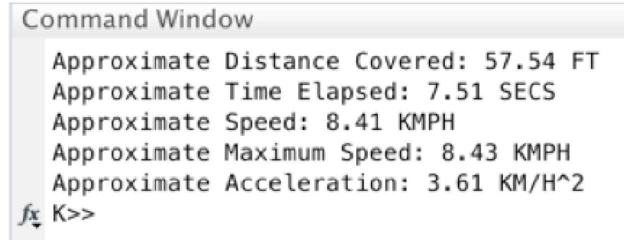

Figure 8. Statistics

The final resultant court coverage map and heat map is given as follows. Please note, the blue region in the heat map shown in figure 9 indicates that the player has not covered that area at all during that rally.

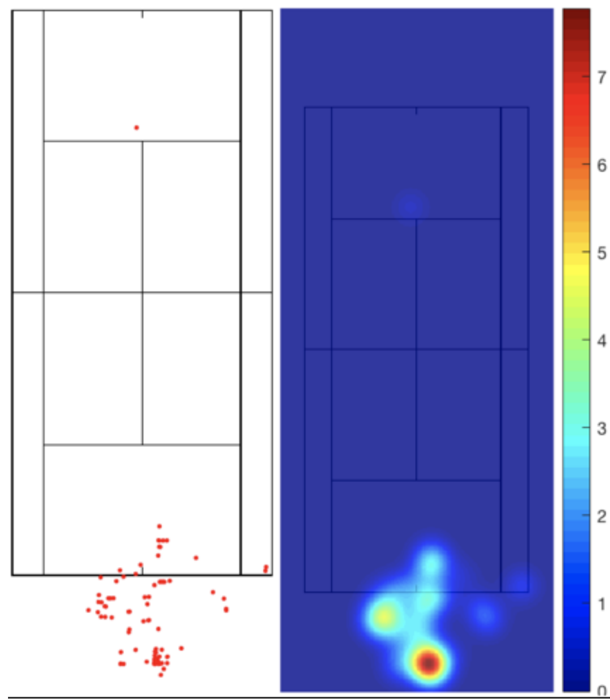

Figure 9. Heat Map

6. Conclusion

As we can observe from the aforementioned results, it is possible to compute statistics of a player over just a single rally. Consequently if the same process is carried out over the course of an entire match, by selectively clipping out areas of the video clip, i.e. moments in between points, games and sets (where the algorithm would fail as it would not be able to capture court points and detect players of set pixel dimensions, it would be possible to realise a certain trend for a given player. Just over the course of a single rally, it is evident, that Rafael Nadal plays mostly on the left-court (which is consistent with the fact that he actually does play most of his points using his stronger forehand).

Such observations along with the calculated statistics are some of the information that could be potentially extracted from implementing this approach into tennis matches videos, meaning that players could come with interesting strategies by analyzing Heat Maps of other players, by just reviewing videos of previous games, as well as obtaining conclusions of how efficient they play.

Some of the limitations of our project has been that due to time constraints, we could not try implementing this approach for the player on the other side of the court as well. We believe that it will be similar (with the only persisting challenge of being able to distinguish the player from the ball boys when he takes a shot deep from outside the baseline). Also, we could not test our algorithm on too much data. It would be interesting to see how it would react to grass courts (as the court lines are not distinctively different in color from the court). We aim to try and test our algorithm and improve it consequently, on more data in the future.

References

[1] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.