A Ball Tracking System for Offline Tennis Videos
Baris David Ekinci, Prof. Dr. Muhittin Gokmen
Department Of Computer Science
Istanbul Technical University
Ayazaga Istanbul
TURKEY

Abstract: - As a result of the increasing demand for higher quality in sports broadcasting, sports video analysis has become a popular research field in computer vision. Innovative replays, content insertion and automated referee systems have become common tools of the trade. Such high level systems require lower level systems capable of extracting basic information automatically. In this paper, drawing on previous work in the field, we present a tennis ball tracking system for offline tennis videos. The system has been on tested on 5 videos extracted from Dubai 2008 and yields promising results.

Key-Words: - Object tracking, Sports video analysis, Background extraction, Candidate feature images, Kalman filtering, Mathematical morphology.

1 Introduction
Due to recent advances in computational technology, research on real time tracking and event classification systems have become more and more popular. As an application, sports video processing has become one of the fundamental requirements for innovative broadcasting. 3D replays, ball speed measurements, event classification, player gesture recognition have become common tools of the trade for broadcasting companies and increasing attention is being redirected to research involving semantic analysis, video cataloguing and highlight extraction.

Information retrieval from sports video is a very challenging research area because it involves the notion of content which is subjective and maybe dependent on context (background, illumination, objects, etc). Object tracking is the most frequently used method in sports videos analysis [1]. A variety of image processing and computer vision techniques have been used to track objects in sports videos [2], [4], [5], [7] while trajectory discrimination remains as one of the most popular [3], [6], [8]. Some tracking techniques use multiple camera models and estimate the 3-dimensional trajectory of objects while others concentrate on single camera shots and are limited to 2-dimensional models [1].

In this study, tennis videos are studied to analyze the motion of the tennis ball. Searching for the ball in a tennis video is not a straightforward subject: The speed of the ball can cause shape differences, it can be over the line and may not be seen at all, or may be occluded by players. Due to the loss of data in video compression or its velocity, color of the ball may be mixed with the background color.

All these difficulties cause problems if direct detection of the ball is concerned. Therefore, methods using different stages to separate moving objects from the static background have been commonly preferred over direct search. Our proposed methodology uses such different levels to track the tennis ball in videos. When considered separately, these levels have been analyzed by various authors in the literature. Common steps for such studies consist of foreground separation or motion detection, ball extraction, and tracking [1].

Foreground separation is not usually done by first calculating the background and then subtracting it from a given frame. Instead, pixel wise temporal differencing is generally used to detect motion [9]. However in this study, a fixed camera is assumed; therefore, the motion on the foreground can be easily separated from the static background. Our method for foreground extraction is similar to [10]. During ball candidate generation, the most widely used method is to filter objects according to their appearance based properties [3], [4]. However, due to various reasons, other objects which have similar properties are usually classified as potential balls and candidate generation results in numerous false balls. We use a trajectory discrimination procedure to eliminate the false positives. Finally, the tracking step realizes the Kalman filtering as usual [4], [11].

We propose a basic framework to be used in domains such as sports video broadcasting. In this framework, the ball tracking problem is split into four steps: (1) Landmark detection and background extraction, (2) ball candidate generation, (3) candidate discrimination, (4) tracking. Each step provides useful information for the next one.
This paper is organized as follows: In section 2, the details about the framework are given. Section 3 represents some experimental results and in Section 4, some concluding remarks are provided.

2 A Ball Tracking System

The presented system assumes a fixed camera that sees the field from behind a player. Although the system is developed for offline videos, it is not difficult to extend it to a real time tracking system. The following will present in detail each of the steps constituting the system.

2.1 Background Extraction and Landmark Detection

In order to come up with a robust foreground extractor, a solid definition of background in the domain of tennis videos that will be analyzed is needed. Since the goal is to track the ball, and by definition, in tennis videos the ball and the players are considered to be almost constantly moving over time, the background is defined as the least moving/changing regions in the totality of the video sequence in question. In other words, foreground objects won't be staying in the same positions for long and the background can be thought of as the "mean image" of the sequence. Hence, a median filter is used over the time dimension of the video. As shown in Fig.1, the median filter results in static background image which can be used to extract the dynamic foreground.

![Original Frame](image1)(a) Original Frame ![Background Image](image2)(b) Background

Fig.1 Background extraction by median filtering

The resulting background image cannot be used directly for landmark (court lines) detection since it carries useless color information. Court lines have some common features that can be exploited for easy calculations. They are distinctively white and linear in all perspectives. Next step is to map the background image into HSV space to utilize the white color. V channel represents the illumination intensities and therefore can be used to distinguish court lines. Top-Hat morphological operator is used on the V channel to get areas brighter than their neighboring. Top-Hat formulation can be given as subtracting Opening of a signal from itself:

\[ \gamma_B(f) = \delta_B(\theta_B(f)) \]  

where \( f \) is the input signal. Dilation \( \delta_B \) and Erosion \( \theta_B \) operations are defined as

\[ \delta_B(f)(x, y) = \bigvee_{(r,s) \in B} f(x - r, y - s) \]  

\[ \theta_B(f)(x, y) = \bigvee_{(r,s) \in B} f(x + r, y + s) \]

B is the selected structuring element to be used during morphological operations.

After the Top-Hat operation, thresholding and finally erosion morphological operations are applied to get noisy court lines. Since we have the general template of the court with exact proportions, it is easy to reconstruct the clean and distinct court lines. Fig. 2 shows the result of each step mentioned.

![V Channel](image3)(a) V Channel ![After Top-Hat](image4)(b) After Top-Hat
![Thresholding](image5)(c) Thresholding ![Model Fitting](image6)(d) Model Fitting

Fig. 2 Landmark detection

2.2 Ball Candidate Generation

Our ball candidate generation method starts by obtaining the foreground objects. The background image is subtracted from each frame of the given video. Just in case the background image contains noise (segments of foreground objects), temporal pixel-wise difference is applied to the resulting frames and all static fields are removed. The rest of the candidate generation process follows the method proposed in [3] in which all foreground objects are filtered according to their shape, size and color. The size of an object is determined by the
total pixels it has according to its 8-connectedness, or simply the size of its bounding box, whereas its shape is determined by the height/width ratio of its bounding box. To filter objects according to their color, each frame is mapped to HSV space and the H channel is used.

The candidate generation process usually results in a frame with the ball and some false positives. Due to noise in measurements, loss of detail in video compression, camera shake, or simply occlusion by players, in some frames even the ball may be filtered and in this situation no candidate will be found. To eliminate false positives and fill in missing data, candidate generation and tracking will be applied.

2.3 Ball Candidate Discrimination

To efficiently represent ball candidates, Candidate Feature Images, a special data structure has been proposed in [3]. A Candidate Feature Image (CFI), also called X or Y-distribution image in [4], contains information about candidates’ positions over frames. To find best candidates for each frame, a discrimination algorithm which keeps track of player positions has been proposed in [3], and a curve-fitting algorithm for volleyball sequences has been proposed in [4]. Since the fixed camera case is simpler (as tilt, pan and zoom are not assumed for the camera) than the situations proposed in [3] and [4], our system utilizes a simplistic method which doesn’t necessitate player position information. The candidates from a frame in the CFI are linked to the nearest neighbor in the next frame; the outliers are thus eliminated from the CFI. Fig.3 and Fig.4 shows the positions of the candidates for each frame. As seen in Fig.4, for each frame system offers only one ball position after discrimination.

2.4 Tracking

The last step of the ball extraction process is tracking. It is obvious that single candidates extracted from CFIs cannot be relied on. To fill in missing information and remove the effects of outliers, two Kalman filters (on the x and y axis) are applied to the trajectory obtained from the discriminated CFI.

The reason for choosing 2 1D filters over a single 2D filter is the sudden change of direction of the ball caused by hitting by players. This challenge is solved in [3] by fixing the ball position to the hitting players position, and in [11] by using two dynamic models for the 2D Kalman filter. The usage of 2 1D filters eliminates the need of two dynamic models and knowledge of player positions.

3 Experimental Results

To evaluate the performance of the tennis ball tracking system, 5 video sequences have been extracted from the Dubai 2008 Open quarter final match between E. Dementieva and A. Ivanovic.

The sequences consist of a total of 1019 frames over approximately 40 seconds. All videos have a bit rate of 48 kb/s and a frame rate of 25 frames/s. The standard deviations of observation and process noise are considered to be both 1 pixel. The reason for this choice is the confidence associated with the background extraction technique. Since the camera is assumed fixed and the estimated background is sufficient, we can assume the process noise will be minimal.

The results obtained are encouraging: a minimum success rate of 95% is achieved in all videos, except for one video which starts from the middle of a game, starting with a couple of object-undistinguishable frames (i.e. a frame in which the ball cannot be seen). The relatively high percentage of observed ball positions (in approximately %80 of the frames the tennis ball is observed, and the remaining 20% of trajectories are estimated) makes the system converge to correct positions quite easily.

<table>
<thead>
<tr>
<th>Video</th>
<th>Frames</th>
<th>Detected</th>
<th>Tracked</th>
<th>Detected Frames</th>
<th>Missed</th>
<th>Success Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>179</td>
<td>318</td>
<td>69</td>
<td>84.13%</td>
<td>9</td>
<td>97.62%</td>
</tr>
<tr>
<td>2</td>
<td>270</td>
<td>218</td>
<td>52</td>
<td>80.74%</td>
<td>8</td>
<td>97.04%</td>
</tr>
<tr>
<td>3</td>
<td>170</td>
<td>99</td>
<td>31</td>
<td>55.71%</td>
<td>3</td>
<td>95.71%</td>
</tr>
<tr>
<td>4</td>
<td>120</td>
<td>90</td>
<td>50</td>
<td>75.00%</td>
<td>5</td>
<td>90.81%</td>
</tr>
<tr>
<td>5</td>
<td>181</td>
<td>118</td>
<td>67</td>
<td>65.19%</td>
<td>24</td>
<td>86.74%</td>
</tr>
</tbody>
</table>

Table 1 Experimental Results
4 Conclusion

A ball tracking system for offline broadcast tennis videos has been presented. An efficient background extraction technique is used to segment moving objects, and multiple visual clues have been used to segment the ball among the foreground. These visual clues include the color, the size and the shape of the ball. The ball candidates are represented in a special data type named Candidate Feature Image. A kalman filter is used to track the ball over frames and fill in missing information in the CFIs. Contrary to what has been used before, instead of using a 2-d kalman filter with 2 different dynamic models which are switched according to the distance of the ball to player locations, this work proposes using 2 separate 1-d filters over frames, hence knowledge of player locations is no longer needed.

References: