AUTOMATIC TRACKING METHOD FOR SPORTS VIDEO ANALYSIS

Jungong Han[a], Dirk Farin[a], Peter H.N. de With[ab], Weilun Lao[a]

[a]Eindhoven University of Technology P.O.Box 513, 5600MB Eindhoven
[b]LogicaCMG, TSE-2, P.O.Box 7089, 5605JB Eindhoven, The Netherlands
jg.han@tue.nl

This paper addresses the implementation of an advanced visual-based tennis sports video analysis system. Firstly, a white pixel-fraction based method is used to detect the significant frames that include the tennis court. In addition, we employ the temporal correlation between two consecutive frames to track the court location within a local search area. Furthermore, we propose a player segmentation and tracking algorithm that separately builds background models for the playing field and the area surrounding the field according to their different colors. This enables to provide it as a prediction of the current frame for more efficiency. Results on real tennis video data are presented, demonstrating the validity and performance of the approach.

1. INTRODUCTION

An upcoming trend in multimedia is to create a large database of AV files for movies and music. For this type of databases, the user desires a management system that enables support for quickly searching and retrieving specific files. While automatic analysis of general video sequences is difficult, or even impossible, the identification of a specific type of video image can be achieved, provided that the video structure can be explicitly defined and its model can be constructed. Sports video analysis [1][2][3] is such a kind of application, which is receiving increasing attention, because it appeals to a large audience. This paper presents our ongoing work [1] to create an automated analyzer of tennis video sequences.

To analyze a tennis video at a higher semantic level, it is necessary to know where the court is, where the players are, as well as the relative position between the court and the players. In [2], a court-line detection algorithm is described, using a straight-line detection method and building a court model. However, it is impossible to obtain good results in the case that the court is partly occluded, and a starting position for the search has to be provided. In [4], a Hough transform is

[1]This research is part of the European ITEA Candela project.

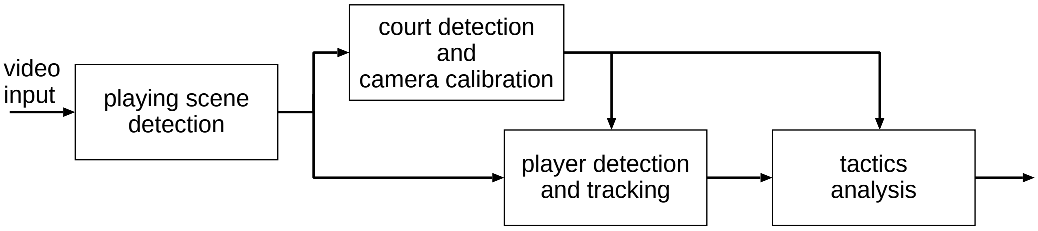

Figure 1: The block diagram of the system.

applied to detect court-lines, which are subsequently used for calibration. However, the approach is based on heuristics to assign the detected lines to lines in the court model. These heuristics are not applicable to the general case. Instead, the research in [5] uses a combinatorial search to establish correspondences between the lines detected with a Hough transform and the court model, which provides a high robustness. However, this method ignores the temporal correlation between two successive frames. The work from [2] introduces a player detection and tracking algorithm, which detects the initial positions of the players in the residue image generated by subtracting successive frames. Afterwards, a Mean Absolute Difference (MAD) matching is employed to track the players over the successive frames. Obviously, with this detection algorithm it is impossible to resolve typical problems of background-based change detection approaches, such as "holes" or "ghosts" created by objects in the background that started to move. Finally, in [6], a change-detection player finding algorithm is proposed that selects the frame of the tennis court without any players present as the background. Unfortunately, in most tennis videos, such frames rarely occur.

This paper describes the pixel-based image and video processing techniques needed to implement a fully automatic tennis video analysis system. Firstly, a novel playing frame detection algorithm will indicate whether the current frame includes court information or not, since only the playing scenes are important for the subsequent processing. Secondly, the court detection algorithm as well as the 3-D court model of [5] are reused in this paper. In our case, we only use it to decide the initial location of the court, then apply the proposed algorithm taking temporal correlation between consecutive frames into account to track the court over the remainder of the video sequence, which largely increases the speed of our system. Furthermore, we present a new change detection-based player segmentation and tracking algorithm. The main contribution of it is that we separately build background models for the space inside the court and outside the court due to their different color features.

The sequel of this paper is arranged as follows. Section 2 introduces the basic framework of our tennis video analysis system. Court and players tracking algorithms are described by Section 3. Section 4 gives the experimental results. Finally, Section 5 concludes this paper.

2. OVERVIEW OF PROPOSED TENNIS VIDEO ANALYSIS SYSTEM

Figure 1 portrays the complete system consisting of five components. First, playing scene detection involves the selection of the tennis playing field sequences out of full sports program including special scenes like e.g. breaks or advertisements. Second, court detection identifies the court location in the scenes and provides specific information, such as size, shape and location. Third, players tracking function determines the position and speed of each player, which are required to derive player tactics. Fourth, the camera calibration deduces a semantic meaning from the position and movements of the players, taking the camera motion into account and computes the player positions in the real-world coordinates. These coordinates are required because the player tracking algorithm will only give the player positions in image coordinates, which are physically meaningless. After the above is implemented, we can analyze tactics and data statistics, such as real running distance and behavior of the players. In this paper, we will focus on the first three modules, and propose some tracking algorithms to resolve practical problems.

3. ALGORITHM DESCRIPTION

3.1. Playing-field Detection Algorithm

In [2], a color-based playing-field detection algorithm is proposed that summarizes the *mean* value in each color space of the four court classes, like carpet, clay, hard and grass, based on statistical analysis using many example frames. The *Euclidean* distance between the color of the current frame and each class of courts is computed, and a class threshold is applied to decide whether it is a playing field or not. Besides a complex procedure for training data, the finding of the threshold is a serious problem, since the mean color of the same court type varies considerably as the function of e.g. the presence of shadows, light conditions and partial occlusion(s).

We have found that the color of the court lines is always white, irrespective of the court type. In addition, the number of the white pixels composing these

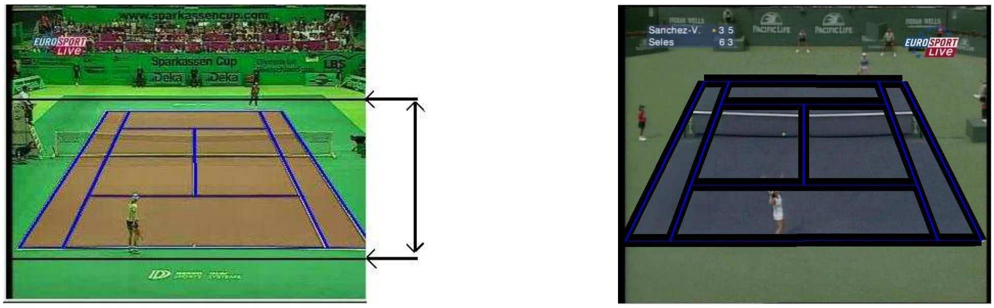

Figure 2: At the left: identification of the local regions for detecting the playing field (black lines, arrows indicate the vertical for the field). At the right: the search area for the court lines (bold black lines structure).

lines is relatively constant over an interval of several hundreds of frames. Based on these two properties, we propose the following algorithm.

1. Initialize with the court detection algorithm of [5], until a court is detected. Get the positions of two baselines and two sidelines, forming a local area as shown in Figure 2 (left), where two black lines represent the vertical boundaries. The horizontal boundary is the same as the picture width.
2. Compute the fraction of the white pixels within the selected area using $F_w = N_w/N_t$, where N_w is the number of white pixels within the area and N_t is the total number of pixels in the area. We employ the technique proposed in [5] to extract white pixels. Its advantage is that white pixels that do not belong to court lines (i.e., the player's white clothing) are rarely marked.
3. Calculate the value of $|F_w(t) - F_w(t-1)|$, where t refers to the current frame and $t-1$ to the previous frame. If it is less than a threshold, this frame is indicated as a frame containing the playing field. The threshold is based on many experiments.

3.2. Court Tracking Algorithm

At the earlier stage, we developed a court detection algorithm [5]. It starts with a model initialization step that locates the court in the image without any user assistance or *a-priori* knowledge about the most probable position. Video pixels are classified as court-line pixels if they pass several tests including color and local texture constraints. A Hough transform is applied to extract line elements, forming a set of court-line candidates. The subsequent combinatorial search establishes correspondences between lines in the input picture and lines from the court model. The algorithm is very robust to occlusion, partial court views, poor lighting, and so on. This technique achieves highly accurate results, but the speed is not sufficient. Therefore, we propose a court tracking algorithm that exploits correlation between two consecutive frames to improve the processing speed.

Tracking of the court is carried out in two steps. The first step is an initialization that captures court information by means of court-line pixel detection, line parameter estimation and model fitting[5] in the first frame containing a playing field. For each detected court line, it is simple to obtain a search area based on the start and end points of the line (see Figure 2). The second step executes the same detection algorithm as the first step, but now within the local search area. The current location of the court is iteratively updated to predict the search area for the coming video frame.

3.3. Player Detection and Tracking Algorithm

In order to track the players over the image sequence, the initial locations of the two players in the picture domain have to be located. For this, a popular moving object segmentation method, called change detection, is used, in which the background is first constructed. Afterwards, the foreground objects are found by comparing the background frame with the current video frame. Background construction is a key issue, as the performance of the change detection technique depends largely on the quality of the background. In most tennis video sequences, a normal frame containing the playing field, mainly includes three parts: the playing field, the area surrounding the playing field and the area for the audience. Normally, the moving area of the players is limited to the field inside the court and partially the surrounding area. Moreover, the color of the playing field is uniform, as is also the case for the surrounding area. Above findings enable us to separately construct background models for the field inside the court and the surrounding area, instead of creating a complete background for the whole image. There are two advantages, when compared to the conventional algorithms. First, a background picture with better quality will be obtained. Second, only spatial information is taken into account as compared to the traditional segmentation algorithms using both spatial and temporal information. Obviously, the motion-estimation module and one frame memory used by the conventional methods are not required for our proposal.

Player Detection in the first frame with a playing field

1. **Construct background for the playing field inside the court.** Up to now, we have obtained the boundaries of the court, and the coordinates of each white pixel of the court lines. We can therefore label the pixels excluding the white pixels within this area as *inside pixels*. After this, the background model for the playing field inside the court is made, in which the intensity of each pixel equals the mean intensity of all *inside pixels*.
2. **Construct background for the area surrounding the playing field.** Predict the moving area for the players outside the field, and label all pixels of this area as *outside pixels*. In order to get a more robust moving area, we predict it using a standard court model constructed in the real world, which is then transformed into the picture domain employing 3-D camera parameters. Once all the *outside pixels* have been extracted, the same averaging technique as mentioned above is applied to construct the background for the area surrounding the playing field.
3. **Produce binary map.** Create a binary map having the same size as the original picture, and initialize all pixels with 255. This map is used later for player position analysis. A residue picture is formed by subtracting the background model from the original picture. The output binary map is obtained as
$$B(x,y) = \begin{cases} 0, & \text{if } |d(x,y)| < Th; \\ 255, & \text{otherwise}, \end{cases} \tag{1}$$
where $B(x,y)$ represents the value of the binary map at a given point (x,y), $d(x,y)$ denotes the corresponding value in the residue picture, and Th is the threshold. Figure 3 shows an example, where the left picture is the original, and the right picture is the produced binary map.
4. **Extract the players**. Exploiting the knowledge of the game court, we select two *search regions* in the binary map, above and below the tennis net-line (see the arrow in the right picture of Figure 3). Subsequently, we scan the Top and the Bottom search regions of the map (above and below the tennis net-line) with two windows of different dimensions (in order to account for perspective differences), referring to the player bodies. For example, we place the player window in the Bottom region W_B at every possible pixel. For each pixel position (the center position of the small window), we count the number of zero values in bottom region W_B enclosed by

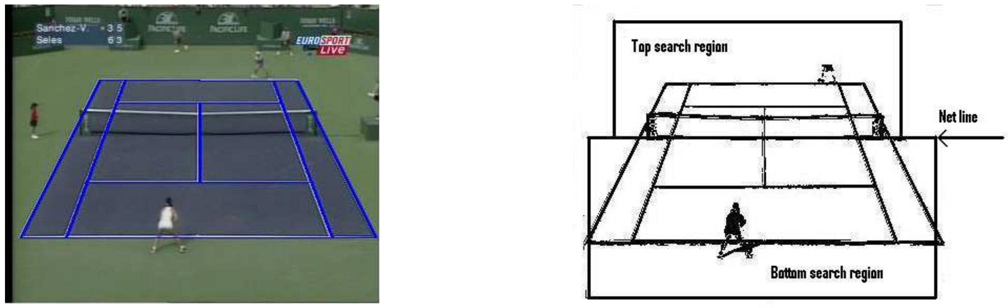

Figure 3: An example of constructing a binary map.

the small window. Then we select max_B, the centroid of the bottom player, and find the related position which maximizes the zero count. Similarly, we obtain the top player position using the same algorithm.

Player tracking in successive frames

When the initial player positions are determined in the first frame with the playing field, we can detect the players in an efficient way for successive frames. We only explore two local windows surrounding the identified small window centers of the previous frame. Within those two local windows, we again search for the positions that maximize the zero counts in W_B and W_T.

4. EXPERIMENTAL RESULTS

We have implemented the proposed tracking algorithms and evaluated the results for various tennis scenes. In our experiments, we followed the functions as described in Section 2 and Figure 1. The results obtained are very promising.

We performed our playing field detection algorithm using three tennis video sequences, consisting of more than 20 minutes and 10,000 frames (25 frames/second). We obtained an impressive 99% detection rate. Furthermore, in order to evaluate the computation efficiency, we tested the court tracking algorithm on a 3 GHz Pentium-4 computer using 720×576 resolution input videos. The speed of the algorithm from [5] is improved with about 20-25%, depending on the complexity of the video frames, while preserving the same accuracy. Finally, we have evaluated our player detection and tracking algorithm. The performance criterion is that at least 70% of the body of the players is included in the detection windows (the black windows shown in the Figure 4). The player detection accuracy is about 98%.

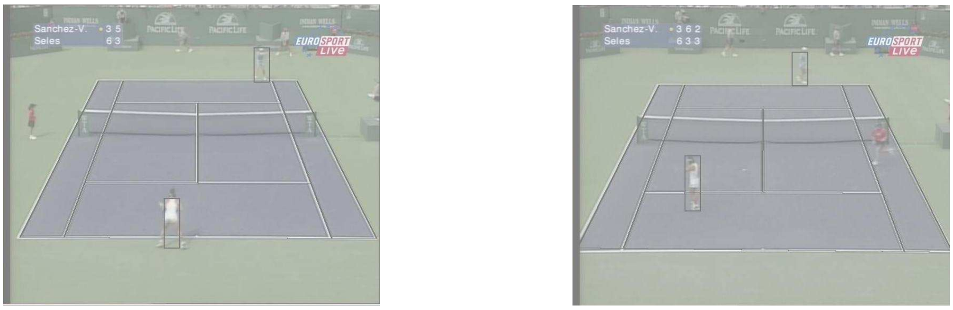

Figure 4: Results of the algorithm for player segmentation and tracking. Left: segmentation result of the first frame. Right: tracking result of the 900th frame.

5. CONCLUSION

We have presented a technique for tennis sports analysis, including the detection of the playing field using court line detection and tracking exploiting temporal correlation, player detection and tracking using a binary map function for two players. The system facilitates analysis of tennis sports video at high level. The system is completely automatic and does not need any human positioning. We envision an integration of our application in a new experimental media server system, which is developed in an international cooperation to enable more extensive experiments.

REFERENCES

[1] A. Ekin, M. Tekalp and R. Mehrotra “Automatic soccer video analysis and summarization”, IEEE Trans. CSVT, Vol 12, pp. 796-807, 2003.

[2] G. Sudhir, C. Lee and K. Jain “Automatic classification of tennis video for high-level content-based retrieval”, Proc. IEEE international workshop on content based access of image and video databases, pp. 81-90, 1998.

[3] Y. Gong, L. Sin, C. Chuan and H. Zhang “Automatic parsing of TV soccer programs”, Proc. ICMCS, pp. 167-174, 1995.

[4] C. Calvo, A. Micarelli and E. Sangineto “Automatic annotation of tennis video sequences”, Proc. DAGM-symposium, Springer, 2002.

[5] D. Farin, S. Krabbe, P.H.N. de With and W. Effelsberg “Robust camera calibration for sports videos using court models”, Proc. SPIE Storage and Retrieval Methods and Applications for Multimedia, volume 5307, pages 80-91, 2004.

[6] T. Bloom and P. Bradley “Player tracking and stroke recognition in tennis video”, Proc. of WDIC, pp. 93-97, 2003.