Ball-path inference based on a combination of audio and video clues in tennis video sequences

BALL-PATH INFERENCE BASED ON A COMBINATION OF AUDIO AND VIDEO CLUES IN TENNIS VIDEO SEQUENCES

[1]*Weilun Lao,* [1]*Jungong Han,* [2]*Peter H. N. de With*

[1]w.lao@tue.nl

[1]Eindhoven University of Technology, P.O. Box 513, 5600MB Eindhoven, The Netherlands

[2]LogicaCMG Netherlands, RTSE-1, P.O. Box 7089, 5605JB Eindhoven, The Netherlands

ABSTRACT

Tennis-sports analysis is attracting much attention in content-analysis research and professional applications. This paper presents a scheme for sports analysis employing an automatic tennis ball-path inference driven by a combination of auditory and visual information. The ball-path inference is implemented for tactics analysis. Since ball tracking remains to be a challenging issue in practice, we use a non-tracking approach for ball-path inference. We propose an effective serving-player detection for achieving an accurate match between a sequence of racket-hit moments and the position of the hitting player in the corresponding video frames. Experimental results have shown that the proposed system can reliably detect the serving-player and classify into different categories, such as left-court/right-court service and front-court/back-court service. Therefore, our system can be utilized for an effective and automatic extraction of various tennis events, performance and tactics analysis with high reliability.

1. INTRODUCTION

Nowadays, sports video (tactics) analysis is becoming increasingly popular and its application appeals to a large consumer audience. It may facilitate practical user requirements involving effective database management, indexing, and quick searching and retrieving of specific contents [1] [2]. In multimedia systems, both audio and video signals can simultaneously contribute to the analysis at the semantic level [3] [4].

The tennis ball-path plays an important role in the tactical analysis of tennis. However, recovering the tennis-ball path remains a challenging task. Approaches for semantic-level analysis based on ball tracking have been explored [5] [6], where a rather accurate trajectory, using tennis-ball detection and tracking, could be obtained. Generally, it is quite difficult to detect and track the ball accurately in practice, especially for a long sequence, because of the following properties of the ball in the image domain.

- The ball size is small.
- The detection is considerably affected by the environment, like lighting, type of playing field, etc..
- The quality of the acquired video sequence is not always sufficiently good to detect the ball.
- The ball deformation is significant, especially at high speeds.

In fact, the accurate trajectory of the ball is not necessary for tactics analysis. Only the ball path is important. Thus we could alternatively utilize a non-ball-tracking approach with the fusion of audio and video clues to obtain the ball-path inference. The approach should also contribute to the tactics analysis in balancing accuracy, robustness to varying circumstances, and the involved computational cost.

The objective of our research is to create an automatic online tennis sports video analyzer at a high semantic level using both auditory and visual information. This paper contributes in two aspects. First, we present a simple and reliable serving-player detection scheme, that classifies the service status and which is driven by fusing audio and video clues about the service. Second, the tennis-ball path is generated for tennis-game tactics analysis without detecting and tracking the ball itself.

In the sequel, we first introduce the architecture of our AV-clue based analysis system in Section 2. The proposed serving-player detection scheme is described in Section 3. The experimental results are summarized in Section 4 and Section 5 concludes the paper.

2. SYSTEM FRAMEWORK

The hierarchical block diagram of our tennis video analysis system is visualized in Figure 1.

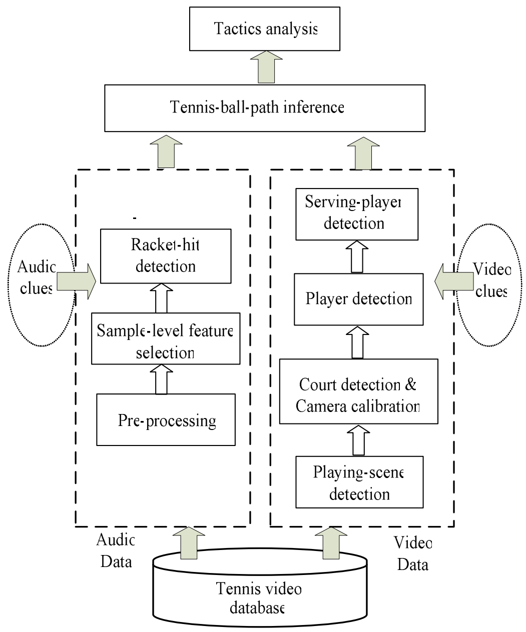

Figure 1. Block diagram of our analysis system.

As a first step, both the video and audio data are extracted from the tennis video sequences, which are fed to two different modules. In the audio-based analysis module, pre-processing filters remove accompanying noises. Subsequently, the audio data is used for sample-level feature detection, employing both time-domain and frequency-domain analysis. Next, a racket-hit detection scheme effectively removes incorrectly detected hits, based on the feature properties and specific game-rules knowledge [7]. In the visual-based analysis module, several techniques of our earlier work are applied to detect the playing scenes [8], track the players [8] and detect the tennis court [8]. Meanwhile, a camera-calibration algorithm [9] is used to bridge the gap between the image domain and the real-world domain. In other words, it transforms the image coordinates to physical positions. After that, a serving-player detection, described in Section 3.2, is activated. Based on the combination of audio and video processing results, we are able to obtain the tennis-ball-path inference. Therefore, the tactics analysis of a tennis game can be implemented accordingly.

3. METHODOLOGY

3.1. Audio-based racket-hit detection

Particular audio events like the racket-hit are important for analysis at the semantic level. We implement the racket-hit detection [6] and obtain the set of racket-hit points $P = \{P_1, P_2, P_3, \ldots, P_n\}$, which has n verified racket-hit points for each rally segment. Afterwards, we classify every service into one of four possible classes: normal service, first-service failure, second-service failure and Ace ball, based on their individual audio properties. The accuracy of audio-based analysis significantly contributes to the other module for visual-based analysis.

3.2. Visual-based serving-player detection

Triggered by the timing of each racket-hit sound, the visual analysis in the image domain is performed. Based on our previous work [8], we obtain the position of each player in each frame. Then their projection on the real-world tennis court model is also calculated. This may lead to further tactical analysis which is described in Section 3.3.

The serving-player detection plays an important role as it contributes to the mapping between a sequence of racket-hit moments and the position of the hitting player in the corresponding video frames. Here we define the player as *front-court player* who is closer to the camera and the other player as *back-court player*. As the service status of two players is correlated, we analyze only the service of the front-court player. The status of the other player can be easily derived from the rules of the tennis game For example, only one player (front-court or back-court) is serving during a tennis game and his service status is either right-court service or left-court service.

Let us now introduce our serving-player detection approach of which the block diagram is portrayed by Figure 2. First, the image frames corresponding to the service event are extracted. From the results of [8], an initial template indicating the position of the front-court player is available. As the 3D real-world position is also calculated, we can classify the serving status into two cases, such as *right-court player* and *left-court player*, according to the position relative to the detected court [8]. The silhouette of the front-court player is produced with a bounding box. This enables to distinguish whether the front-court player is serving or not, based on a specific silhouette property.

Figure 3 shows an example of the difference between a serving posture and a non-serving posture. The feature of the aspect ratio α of the bounding box can distinguish between a serving player and a non-serving player. This aspect ratio α is defined as the ratio between width W and height H of the bounding box, and which is compared to a threshold such that

$$\alpha = \frac{W}{H} = \begin{cases} \le \lambda & \textit{is serving} \\ > \lambda & \textit{is NOT serving} \end{cases} , \qquad (1)$$

where λ indicates an adaptive threshold which is depending on the relative court size. If the aspect ratio α of the bounding box indicates a service of the front-court player, the service status is classified as front-court service. Otherwise, the service status is labeled as back-court service.

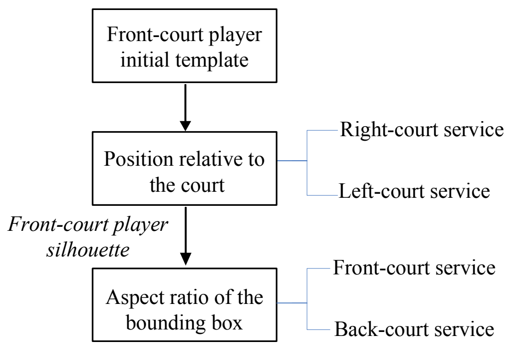

Figure 2. Block diagram of analyzer for serving-player.

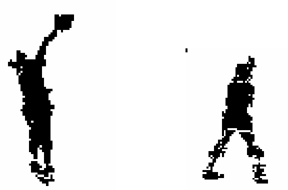

Figure 3. Example of comparing the silhouettes of a serving player (left) and a non-serving player (right).

3.3. Tennis-ball-path inference for tactics analysis

Suppose we obtain a set of n racket-hit points $P = \{P_1, P_2, P_3, \ldots, P_n\}$ from Section 3.1 and a set of its corresponding positions for the front-court player $F = \{F_1(x_{F1}, y_{F1}), F_2(x_{F2}, y_{F2}), \ldots, P_n(x_{Fn}, y_{Fn})\}$ and the back-court player $B = \{B_1(x_{B1}, y_{B1}), B_2(x_{B2}, y_{B2}), \ldots, B_n(x_{Bn}, y_{Bn})\}$. From the serving-player detection and tracking results, we are able to implement the tactics analysis, while no highly accurate ball trajectory is required.

An example of the analysis is shown in Figure 4. In the tennis court model, the circle and the rectangle represent the positions of the front-court player and back-court player when they are making the racket-hit, respectively. The numbers shown in the figure indicate the order of the racket-hit in a particular sequence. The dotted arrow and the solid arrow represent the player trajectory and the tennis-ball-path during the game, respectively. Evidently, this arrangement contributes to the tennis tactics-analysis application. It facilitates the player and/or the coach to analyze and improve the playing performance.

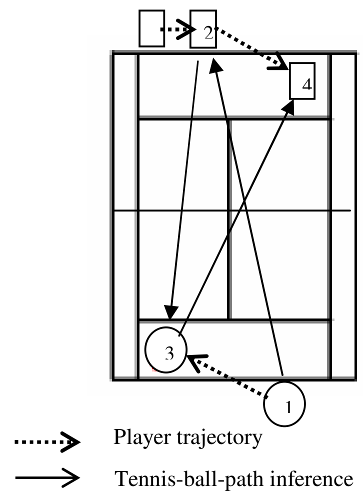

Figure 4. Example of the tennis tactics analysis with indicated ball-path, playing order and player trajectories.

4. EXPERIMENTAL RESULTS

Experiments are conducted to test the performance of our proposed scheme, as described in the previous sections. The database for experiments is composed of two MPEG-2 tennis video clips with a total length of 6 minutes.

Firstly, we verify the proposed serving-player detection scheme, which is described in Section 3.2. The video clips contain 11 actual services in total. Given the accurate racket-hit detection results from audio-based analysis, the accuracy rate of video-based serving-player detection reaches 100% for this (limited) dataset. It should be noted that this high accuracy relies on the audio-based analysis result. We have verified that the performance deteriorates when the audio analysis does not work properly. Employing the algorithm of Section 3.2, we have succeeded to classify the serving status into the four categories as shown in Figure 2 (front/back-court, left/right-court service).

Secondly, for further experiments, we have integrated our detection scheme into a full tennis video analysis system. It runs under the Linux operating system and is programmed in C++. The user interface is shown in Figure 5. The system visualizes the analysis results at three different levels. At the audio level, racket-hit points are detected and further classified into several types of serving and returning. The game status is detected, involving rally, score and break. At the video level, the players and tennis court are tracked and highlighted. At the AV-fusion level, the serving-player status is detected. Moreover, a tactics analyzer is implemented, of which example analysis results are depicted in Figure 6. This effectively and accurately simulates a tennis game at a position level, where the tennis-ball-path is drawn in solid lines between the player positions.

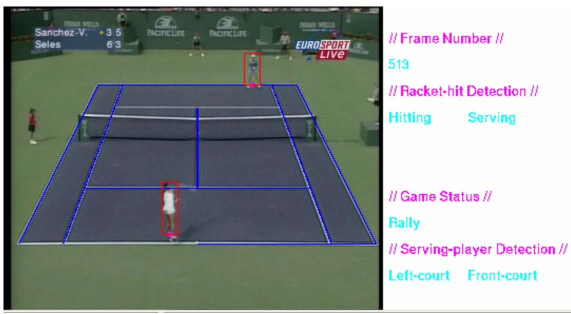

Figure 5. User interface of our tennis analysis system.

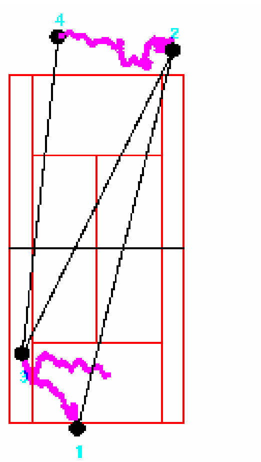

Figure 6. Schematic results of tennis tactics analysis.

5. CONCLUSION

In this paper, we have presented a scheme for automatic tennis-ball-path inference for tennis sports video analysis, based on simultaneously combining audio and video clues. Firstly, an effective serving-player detection algorithm is employed which aims at locating each player in the image domain, corresponding to each racket-hit that is detected by audio-signal analysis. The player detection and tracking is an aid in finding the service by the aspect ratio of the player's silhouette and afterwards accurately classifying the service into several categories such as left-court/right-court service and front-court/back-court service. Secondly, the tennis-ball-path is generated for tennis game tactics analysis without detecting and tracking the ball itself, by projecting players in a real-world model and combining previous analysis with game rules. It was shown that our system can deliver a tactical analysis based on the fusion of video and audio analysis modules. Various functions such as racket-hit detection/classification, game status classification and serving-player detection are also available. The potential application of our system involves event classification, performance analysis and game summarization.

6. REFERENCES

[1] A. Ekin, A. M. Tekalp, R. Mehrotra, "Automatic soccer video analysis and summarization", *IEEE Trans. Circuits. & Systems Video Technol., Vol. 12*, pp. 796-807, 2003.

[2] X. Yu and D. Farin, "Current and emerging topics in sports video processing", *Proc. ICME 2005,* Amsterdam, The Netherlands, July 2005.

[3] M. Xin, L-Y Duan, C-S Xu and Q. Tian, "A fusion scheme of visual and auditory modalities for event detection in sports video", *Proc. ICME 2003*, Maryland, U.S., Vol. 1, pp. 333-336, 2003.

[4] R. Dahyot *et al.*, "Joint audio visual retrieval for tennis broadcasts", *Proc. ICASSP 2003*, Hong Kong, China, Vol.3, pp. 561-564, 2003.

[5] H. Miyamori, "Video annotation for content-based retrieval using human behavior analysis and domain knowledge", *Proc. IEEE Int. Conf. Autom. Face & Gesture Recognition*, Grenoble, France, 2000, pp. 320-325, 2000.

[6] X. Yu, C-H. Sim, J. R. Wang and L. F Cheong, "A trajectory-based ball detection and tracking algorithm in broadcast tennis video", *Proc. ICIP 2004,* Singapore, pp. 1049-1052, 2004.

[7] W. Lao, J. Han, P.H.N.de With, "Automatic sports analysis using audio clues and context knowledge", submitted to EuroIMSA 2006, Austria, 2006.

[8] J. Han, D. Farin, P. H. N. de With, and W. Lao, "Automatic tracking method for sports video analysis", *Proc. Int. Symp.* Inform. Theory in the Benelux, Brussels, Belgium, pp. 309-316, May 2005.

[9] D. Farin, J. Han and P.H.N. de With, "Fast camera-calibration for the analysis of sports sequences", *Proc. ICME 2005*, Amsterdam, The Netherlands, July 2005.