An Automatic Analyzer for Sports Video Databases Using Visual Cues and Real-World Modeling*

Jungong Han[1], Dirk Farin[1], Peter H.N. de With[1, 2] and Weilun Lao[1]
[1] Eindhoven University of Technology P.O.Box 513, 5600MB Eindhoven
[2]LogicaCMG, RTSE, P.O.Box 7089, 5605JB Eindhoven, The Netherlands

***Abstract*—With the advent of hard-disk video recording, video databases gradually emerge for consumer applications. The large capacity of disks requires the need for fast storage and retrieval functions. We propose a semantic analyzer for sports video, which is able to automatically extract and analyze key events, such as player behavior. The analyzer employs several visual cues and a model for real-world coordinates, so that speed and position of a player can be determined with sufficient accuracy. It consists of four processing steps: (1) playing event detection, (2) court and player segmentation, as well as a 3-D camera model, (3) player tracking, and (4) event-based high-level analysis exploiting visual cues extracted in the real-world. We show attractive experimental results remarking the system efficiency and classification skills.**

I. INTRODUCTION

In consumer video applications, sports video attracts a large audience and recording of such programs is popular. However, the enormous amount of AV footage produced and the use of large capacity storage media asks for structured storage and retrieval. The more data is stored, the more consumers need support in organizing their databases. Content abstracting based on key events has been considered, but only partially resolves this problem [1][2].

Significant research has been devoted to the event-based sports video analysis in the past few years. In [2], visual cues are utilized like segmented players, shape descriptions and the game court structure to analyze a sports video. However, the visual cues extracted in the image domain do not exactly reflect the real activity of a player. In [3], both visual and audio features are combined, such as color, player position, crowd sound, vocal announcements and ball hits, to detect specific events in various sports games. Obviously, audio analysis improves the system analysis performance, but it also increases the complexity. In this paper, we increase the level of video analysis to player and scene behavior, in order to come to a much better performance than key-frame extraction.

As an example, we focus on a tennis sports video analysis system. The system exploits several simple but efficient real-world visual cues to classify and extract the key events and scenes. Furthermore, unlike existing proposals only extracting events, we also attempt to study the contextual events, to give a further semantic analysis of the game (e.g. playing style). Our analyzer achieves (near) real-time speed with promising results. Most of the techniques can be applied to other sports types as well.

II. SYSTEM DESCRIPTION

For tennis sequences, it can be observed that, despite the typical broadcasted program length of a few hours, only parts of it contain the real actions of the underlying game. It can be useful to model the game in order to facilitate the scene analysis. This principle is elaborated further in the following system diagram.

A. Playing Event Detection

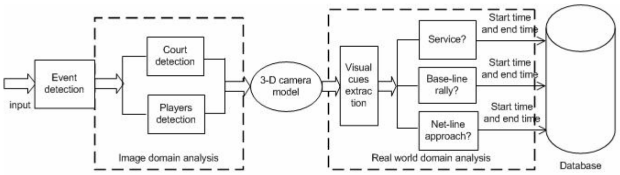


Fig. 1 Framework of the proposed system

The first module of our system shown in the Fig. 1 is playing event detection, whose purpose is to select the tennis playing filed sequences out of a full sports program. We propose a simple but efficient approach [4], that only identifies the white pixels of court lines and distinguishes the difference between the number of white pixel of two consecutive frames. We used this metric, because we found that the color of the court line is always white, irrespective of the court type, and the number of the white pixels composing court lines is relatively constant over an interval of several hundreds of frames. Compared with conventional methods [5] based on *mean* color value, our technique is more efficient and abridges a complex procedure for training data.

B. Image domain analysis

The second step is to segment and track key objects in the image domain, such as court (playing field) and players. The system applies earlier results of our work for court detection [6] and player segmentation [4]. The former algorithm detects white court lines and then fits with a standard tennis court for finding which line in the image corresponds to which line in the real-world model. This method is very robust to occlusion, partial court view and poor lighting conditions, which is better than the methods in [2][5]. The player segmentation employs a change detection-based object segmentation method, and summarizes several effective visual properties in the tennis video (e.g. uniform court color) to build a high quality background, thereby achieving more accurate segmentation results.

C. Real-world player tracking

Unlike conventional algorithms [7][8] that track moving objects in the image domain, we track players in the real-world domain. From [7], they have adopted the Double Exponential Smoothing (DES) operator to track moving persons, which runs approximately 135 times faster than the popular Kalman filter-based predictive tracking algorithm [8] with equivalent prediction performance. Equation (1) defines the double exponential smoothing operator by

$$\begin{cases} s_t = \alpha y_t + (1-\alpha)(s_{t-1} + b_{t-1}), & 0 \le \alpha \le 1; \\ b_t = \gamma(s_t - s_{t-1}) + (1-\gamma)b_{t-1}, & 0 \le \gamma \le 1; \end{cases} \qquad (1)$$

where y_t is observed position value, s_t refers to the output value after smoothing, b_t represents the trend of the player position. α and γ are two weighting parameters controlling motion smoothness, which are usually obtained by non-linear optimization techniques.

* This research is part of the European ITEA Candela project

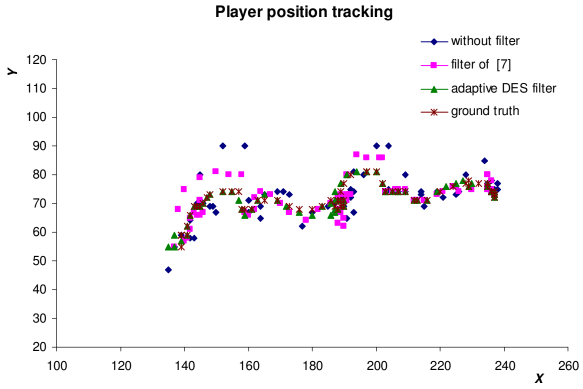


Fig. 2 Player position tracking, using various filter techniques. *X* and *Y* refer to an image domain coordinate system (we track positions in the real-world domain, then transform them to the image domain). The unit of distance is expressed in pixel.

We have found that an adaptive adjustment of filter parameters gives better tracking results. The basic idea is that if the speed exceeds a boundary or a significant speed change occurs suddenly, the probability of false segmentation is increased (adapting the values of α and γ makes the current predictive results more rely on the past trend, as the running speed of a tennis player is normally between 2~7 m/s). Fig. 2 shows examples of player position tracking with various filters smoothing the position coordinates, where the results of our adaptive DES compares favorably to the manually extracted ground truth data.

D. High-level semantic analysis

For scene analysis at high level, we first model the characteristics of each important sub-event such as service, base-line rally and net approach, described from the camera viewpoint. Secondly, based on those models, we try to extract the occurrence of sub-events. Thirdly, we attempt to classify the game type, taking the correlations between each sub-event into account.

1. Visual cues of each sub-event

Service: service is normally started at the beginning of a playing event, where two players are standing on the opposite half court, i.e., one is at the left part of the court, and the other is at the right part. In addition, the receiving player has limited movement during the service.

Base-line rally: the base-line rally is usually after the service, where two players are moving along the base-lines with relative smooth speed, that is, there is no drastic speed change.

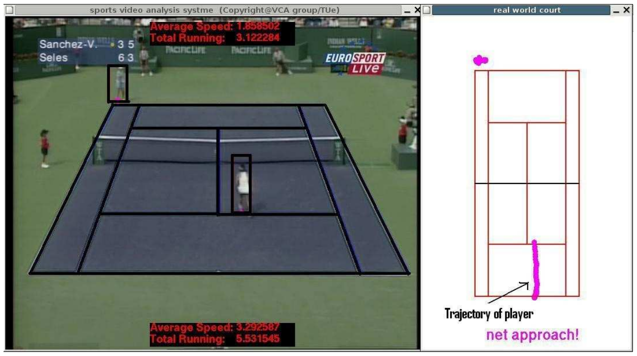
Net approach: this is one of the highlight parts of a game, in which standard cues are large speed change close to the net lines.

2. Extract sub-events based on visual cues

With the visual cues mentioned above, it is easy to decide what kind of sub-events the current frame belongs to. Afterwards, we can detect the start time and end time of each event using a temporal filter.

3. Abstract the game

Our system not only extracts some important sub-events, but also intends to summarize the game, making use of correlations between sub-events. For instance, if there is a service event without a base-line rally that directly changes to a "no event", it is reasonable to conclude that such case might be an Ace ball or a double service fault. Furthermore, it is easy to calculate how many net approaches each player carried out during a match. Based on this, the player with more net approaches is classified as more aggressive.



Average Speed: 1.858502
Total Running: 3.122284

Average Speed: 3.292587
Total Running: 5.531545

Fig. 3 Results of our analyzer. The left image shows the detected court and players, as well as average speed of each player. The right is a real-word court model, where the trajectory of each player is visualized, also showing a net approach.

Table. 1 Classification results

	Detect	Correct	Miss	False
Service	21	21	2	0
Base-line Rally	16	14	0	2
Net approach	8	7	0	1

III. EXPERIMENTS AND CONCLUSIONS

We performed our tennis analyzer using more than 20 minutes of various tennis sequences (25 frames/s). It achieves a 99% detection rate on playing events, 98% detection rate on players, where the criterion is that at least 70% of the body of the player is included in the detection window. Furthermore, the high-level sub-event extraction rate of this system is around 90%, which is listed in Table 1. Fig. 3 shows an example of a tennis video analysis results. It is efficient, achieving a real-time or near real-time performance (2~3 frames/second for 720*576 resolution, and 5~7 frames/second for 320*240 resolution, with a P4-3GHz PC).

Our tennis analysis system was integrated into a networked home multimedia application as a video content analysis unit, which was demonstrated live at an international multimedia conference.

REFERENCES

[1] B. Li, J. Errico, H. Pan, and I. Sezan, "Bridging the semantic gap in sports," *SPIE on Storage and Retrieval for Media Databases 2003*, pp. 314-326, 2003.

[2] C. Calvo, A. Micarelli, and E. Sangineto, "Automatic annotation of tennis video sequences," *Proceeding of the 24th DAGM symposium on pattern recognition,* .London, UK, pp. 540-547, 2002.

[3] N. Babaguchi, Y. Kawai, and T. Kitahashi, "Event based indexing of broadcasted sports video by intermodal collaboration," *IEEE Trans. Multimedia,* vol. 4, pp. 68-75, Mar, 2002.

[4] J. Han, D. Farin, P.H.N. de With, and W. Lao, "Automatic tracking method for sports video analysis," *Proc. Symposium on information theory in the Benelux*, Brussels, Belgium, pp. 309-316, May, 2005.

[5] 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.

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

[7] J. J. Laviola Jr, "An experiment comparing double exponential smoothing and Kalman filter-based predictive tracking algorithms," *Proceeding of the IEEE virtual reality 2003,* pp. 283-284, 2003.

[8] W. Greg and G. Bishop, "An introduction to the Kalman filter," Technical report TR 95-041, Department of computer science, University of north Carolina at chapel hill, 1995.