Computer Vision and Image Understanding 000 (2017) 1–12

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

A technology platform for automatic high-level tennis game analysis

Vito Renò[a,*], Nicola Mosca[a], Massimiliano Nitti[a], Tiziana D'Orazio[a], Cataldo Guaragnella[b], Donato Campagnoli[c], Andrea Prati[d], Ettore Stella[a]

[a] National Research Council of Italy, Institute of Intelligent Systems for Automation, via Amendola, 122 D/O, 70126 Bari (BA), Italy
[b] Politecnico di Bari, Dipartimento di Ingegneria Elettrica e dell'Informazione, via Orabona, 4, 70125 Bari (BA), Italy
[c] Mas-Tech srl, via Cantone, 96, 41032 Cavezzo (MO), Italy
[d] University of Parma, Department of Information Engineering, Parco Area delle Scienze, 181/a, 43124 Parma (PR), Italy

ARTICLE INFO

Article history:
Received 15 March 2016
Revised 7 December 2016
Accepted 2 January 2017
Available online xxx

Keywords:
Synchronized cameras platform
Ball and player tracking
Trajectory analysis
Semantic interaction interpretation

ABSTRACT

Sports video research is a popular topic that has been applied to many prominent sports for a large spectrum of applications. In this paper we introduce a technology platform which has been developed for the tennis context, able to extract action sequences and provide support to coaches for players performance analysis during training and official matches. The system consists of an hardware architecture, devised to acquire data in the tennis context and for the specific domain requirements, and a number of processing modules which are able to track both the ball and the players, to extract semantic information from their interactions and automatically annotate video sequences. The aim of this paper is to demonstrate that the proposed combination of hardware and software modules is able to extract 3D ball trajectories robust enough to evaluate ball changes of direction recognizing serves, strokes and bounces. Starting from these information, a finite state machine based decision process can be employed to evaluate the score of each action of the game. The entire platform has been tested in real experiments during both training sessions and matches, and results show that automatic annotation of key events along with 3D positions and scores can be used to support coaches in the extraction of valuable information about players intentions and behaviours.

1. Introduction

Sport video analysis has attracted much research in the last decade and a wide spectrum of possible applications have been considered such as verification of referee decision, tactics analysis, automatic highlight identification, video annotation and browsing, content based video compression, automatic summarization of play, player and team statistic evaluation and many others (Conaire et al., 2009; D'Orazio and Leo, 2010; D'Orazio et al., 2009b; Hughes and Franks, 2004; Kapela et al., 2015; Lai et al., 2011).

Sports analysis can be either performed using broadcast videos or monocular/multi camera videos acquired by dedicated and optimized cameras (Archana and Geetha, 2015; D'Orazio et al., 2009a; Leo et al., 2008). One of the primary advantages available when using images edited for broadcast is the ability to have a potential access to an extremely large set of data that have been recorded in the past for the masses. Video editing performed on broadcasted video, combined with the emphasis posed in video replays, can also be exploited to aid the recognition of important sport events. At the same time, broadcast videos are generally made available for general audience fruition and are not necessarily acquired under the best conditions for performing automatic video processing to analyze performances of teams and players by means of machine vision algorithms. Moreover broadcast video introduce an inherent bias in the recognition of particular events at the detriment of others. Dedicated multi-camera systems can therefore prove more effective in performing automated and unbiased analysis tasks.

In this paper we introduce a technology platform for the segmentation and the analysis of tennis video sequences by means of four synchronized cameras for data acquisition and some processing modules for 3D trajectories reconstruction of ball and players, automatic semantic analysis of key events, indexing and match scoring.

1.1. Related works

Sports analysis can provide a complete survey of sport events to interested parties. This kind of systems produces objective feedback helping players and coaches to improve performance in a field that is competitive by nature. For this reason, several

* Corresponding author
E-mail address: reno@ba.issia.cnr.it (V. Renò).

http://dx.doi.org/10.1016/j.cviu.2017.01.002
1077-3142/

commercial solutions are available, sometimes addressing analysis in more than one sport discipline. Most of them provide only support for manual annotations of video sequences. Manual annotations can either be done off-line or in real-time. Dartfish Video Analysis Software (Dartfish, 2016) and Sportscode Performance Analysis (Avenir Sports, 2016) are examples of commercial systems in which video sequences are manually annotated off-line on desktop-class computers with the latter being used by important football clubs (Kokaram et al., 2006). Match Analysis (Match Analysis, 2016) is a further example where manual annotations are created, although in this case the operation is remotely outsourced to other companies. TenniVis is a tennis match visualization system that relies entirely on data such as score, point outcomes, point lengths, service information, that can be easily collected by a human operator watching videos that are captured by one consumer-level camera (Polk et al., 2014). Other systems (Performa Sports, 2016; Protracker Tennis, 2015) offer support for smart-phones and tablets, while still requiring manual annotations.

Few systems try to provide a solution to sport analysis without requiring human supervision. A recent trend is represented by the acquisition of data directly from the player using wearable devices in several sports in general (Ermes et al., 2008; Strohrmann et al., 2011) and in tennis in particular (Ahmadi et al., 2009; Connaghan et al., 2011). However, intrusive systems can be either sensible to signal collisions and interferences for operating and communicating in real-time (Chen and Pomalaza-Ráez, 2009), or are limited to off-line processing (Bächlin et al., 2009; Ghasemzadeh et al., 2009). Additionally they are rarely accepted by players as they have to be small enough to be comfortable and not perceived as an obstacle to their movements and performance (Chi, 2005). Non intrusive solutions are based on broadcast cameras or dedicated cameras placed around the game court and use computer vision techniques to process the acquired videos. ProZone (Valter et al., 2006) provides automatic video analysis for soccer and rugby. This system is based on the automatic processing of NTSC/PAL video. The system can operate in almost every professional match broadcasted in TV. Human intervention is sometimes required to correct errors done by the system. TennisSense (Conaire et al., 2009) has been developed to use custom-installed cameras, optimized for automatic processing. The system has been designed and developed by Dublin City University in partnership with Tennis Ireland, the Irish tennis governing body, using a UbiSense spatial localization system and requiring the installation of nine IP cameras with pan, tilt and zoom capabilities, surrounding the instrumented tennis court. Cameras position and setup are optimized to cover specific areas and perform specific tasks. Ball and players tracking is therefore performed synchronizing and fusing these data streams. A system, operating in real time and aimed at enhancing broadcasts as well as coaching activities, is proposed in Pingali et al. (1999); 2000), where computed motion trajectories, along with compressed video streams, are stored in a database system. The system proposed in Pingali et al. (1999) also provides a way to customize information to be shown using a proprietary Application Programming Interface (API).

Other works focus their attention on a more limited set of topics, such as stroke detection or ball trajectories reconstruction. In Bloom and Bradley (2003) strokes are detected and recognized through player tracking and skeletonization, although under restrictive assumptions. Ball trajectory is the focus of the work described in Yu et al. (2003), that is performed on soccer matches using broadcast video. Novel in this work is their focus on recognizing the ball through the evaluation of the followed trajectory rather than its low-level visual features. The ability of discerning event cues starting from the evaluation of ball trajectories is the focus of the work (Yan et al., 2005) on broadcasted tennis matches, enabling therefore automatic annotation of broadcasted videos. Issues on the reconstruction of ball trajectories are also common in table tennis games, with the aggravating problem imposed by frequent occlusions between ball and racquet. The paper Tamaki and Saito (2013) addresses this challenge through the evaluation of trajectory planes. Misdetection and abrupt changes of ball trajectories are addressed in Yan et al. (2006) using a layered data association scheme. Last but not least, ball tracking can be done in 3D using a physics-based approach (as in Poliakov et al., 2010), when sports events are acquired using multiple synchronized views.

In this paper we propose an innovative approach for event recognition such as strokes, bounces and serves, based on the analysis of the reconstructed 3D ball trajectory which can be used for automatic annotations of video sequences and high level semantic analysis. The extracted action sequences with the associated data can support coaches for the evaluation of game tactics and for improving players performance.

1.2. The proposed system

The proposed system consists of a dedicated hardware setup (cameras and computer) and a number of software modules for the automatic processing of the recorded video sequences. The aim is to records tennis video sequences and performs the segmentation and the analysis of significant tennis actions in order to support coaches in the evaluation of tennis players performance during training sessions or official matches.

We propose the use of dedicated cameras in order to collect data that cover all the court and are able to observe simultaneously the positions of players and ball during actions. Broadcast cameras (which commonly show a single point of view of the match) are not suitable for this kind of tasks first because 3D reconstruction of the ball trajectory is necessary to evaluate events, and also because positions of the two teams and the ball in the court are necessary to evaluate tactics and performance. Moreover, broadcast camera videos are often chosen for entertainment purposes then they are not suitable to record all the events necessary for automatic game and players evaluation.

In this paper we propose four synchronized cameras placed on the corners of the court and connected to a central node. They are provided with suitable algorithms for processing the acquired videos. The system reconstructs the 3D ball trajectory and recognizes key events by the concatenation of simple action parts which concern ball rebounds, shots or faults. The main contribution of this paper is, from one side, the system architecture, in terms of number and positions of dedicated cameras, frame rate and resolutions which respect the constraints imposed by the tennis domain. On the other side a number of processing modules has been implemented to perform low level image processing and high level semantic interpretation and to recognize key events automatically assigning a score. Finally, large attention has been put on designing a technology platform that can effectively support coaches with relatively low cost equipments. The results demonstrate that the proposed system is able to effectively reconstruct 3D ball trajectories, recognize serves, strokes and bounces and make a decision about score assignment for each action. Moreover, coaches can perform strategic queries to analyze players intentions, behaviours and performance using a combination of both 3D data and key events annotations.

The remainder of this paper is organized as follows. Section 2 is devoted to describe the proposed system with an emphasis on its hardware architecture and the processing modules. Section 3 focus on the processing algorithms from the low level step to the final decision process that assesses a score. Experimental results are reported in Section 4. Conclusions and further work are finally presented in Section 5.

Please cite this article as: V. Renò et al., A technology platform for automatic high-level tennis game analysis, Computer Vision and Image Understanding (2017), http://dx.doi.org/10.1016/j.cviu.2017.01.002