A technology platform for automatic high-level tennis game analysis

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

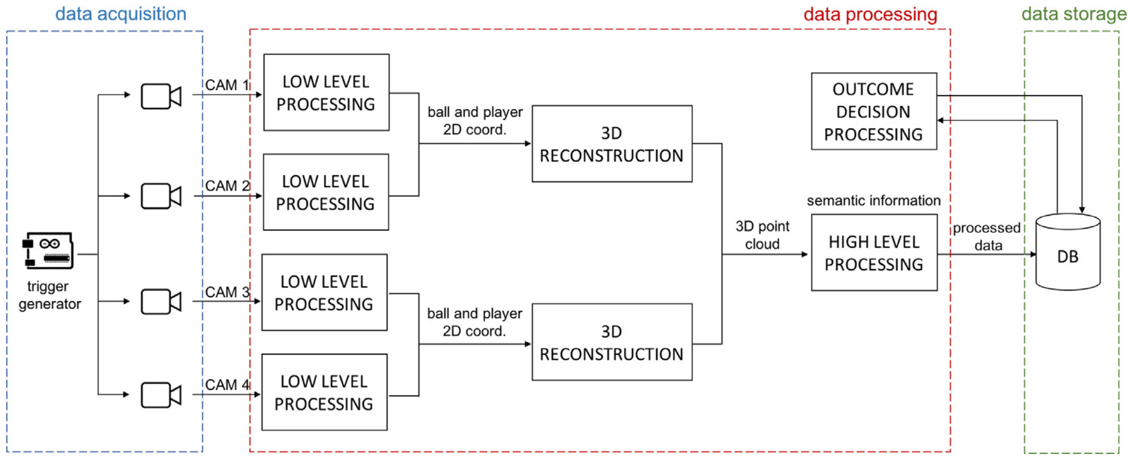

Fig. 1. Block diagram of the system.

2. System overview

The proposed system is primarily designed to address coaching needs. For this reason, it is designed to operate in structured environments, typically indoor, although it can operate outdoor as well. The system can be schematically decomposed in three sub-systems (see Fig. 1): data acquisition, data processing and data storage.

2.1. Data acquisition

In order to design a platform which can operate in every condition, both indoor and outdoor, the most general conditions have been considered to choose the proper equipments and functionalities. Most of the tennis training sessions are performed in indoor environments, generally under shelter structures. Shelter structures create particularly challenging conditions for image processing, since they are made to let sunlight in, therefore with varying illumination from sunny to cloudy as for outdoor condition, and yet allow to switch on artificial lights when sunlight proves insufficient. In addition, artificial lights introduce flickering effects which can greatly modify the quality of the acquired images. For this reason the platform has been designed to operate in the most challenging situation, i.e. indoor environments, but the same equipments can be used or either simplified to operate on outdoor courts.

The proposed architecture makes use of four high-definition cameras that are synchronized using a trigger circuit. The model chosen is an AVT Prosilica GT 1920C. This is a Gigabit camera with a maximum resolution of 1936 × 1456 with a maximum frame rate of 40 fps. However, since the system has been designed to operate also with artificial lights, some horizontal scanlines have been cropped to achieve a fixed frame rate of 50 fps, which is the same of the power lights fluctuations (which operates at 50 Hz in Europe). This frame rate boost essentially removes any flickering issues due to the use of artificial lights and enables the system to operate better during serves, where ball usually travels at its maximum speed. Under these conditions acquired frames have a resolution of 1920 × 1024 pixels. These cameras are also equipped with auto-iris lens control, enabling to stabilize brightness even during extended recording sessions, spanning several hours.

Cameras arrangement requires the use of two pairs of cameras. Each pair is positioned to cover the half-court on their opposite side. The fields of view of a pair of these cameras is shown in Fig. 2 as the blue regions. The intersection of both fields of view is the area in which 3D reconstruction can take place by means of triangulation techniques. A central node, equipped with the trigger generator, synchronizes the four cameras and collects data. The synchronization task has been implemented with the aid of a low cost programmable micro controller, namely an Arduino, that has been set to operate at 50 Hz. A squared wave is then generated accordingly to the selected frequency. This way, every installed camera can acquire an image exactly at the same time, meeting the requirement of each stereo vision system that needs a synchronized pair of images to compute the 3D information. Cameras are connected to the central node by using a Power Over Ethernet (PoE) Gigabit card. The central node is also equipped with dedicated SATA hard disks hosted on a SAS controller, independent from the operating system boot disk, providing storage capabilities that are both sized for extended playing sessions, and shielded from the interference of operating systems tasks. Each camera stream is physically stored on a dedicated partition of a different drive, thus enabling the system to store multiple raw data streams in parallel at the maximum available speed. An example of four synchronized images acquired by the four cameras is reported in Fig. 3.

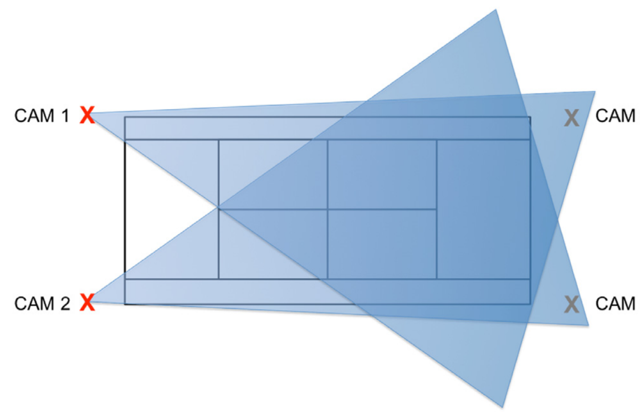

Fig. 2. The figure depicts the position of the four cameras on the corners of the tennis court with X marks. Each pair of cameras can acquire the opposite part of the court. The fields of view of two cameras are highlighted in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. Data processing

The data processing consists of several modules (see Fig. 1): low level processing, 3D reconstruction, high level processing, outcome decision processing.

Fig. 3. Example of synchronized acquisition. Four frames are captured exactly at the same time to make the system capable of observing the whole court from at least two different points of view.

Thanks to the nature of the dedicated installation, using fixed cameras with fixed zoom, low level processing performs moving blobs detection on each camera aided by a background subtraction approach. In this way, low level entities (i.e. ball and player candidates) present in each camera are extracted. Therefore for each camera, a temporal and spatial blob analysis is performed for filtering false candidates due to noise, for recognizing ball and players and remove artifacts due to shadows or net movements.

The resulting ball and players candidates of the corresponding pair of cameras of the same side of the court, are then forwarded to the 3D reconstruction module. For all these candidates, the 3D point cloud is extracted by applying a triangulation approach which exploits known geometrical relationships between cameras, and fixed positions of entities and corners of the court in the real world.

Finally, the 3D point cloud is processed by the high level processing module which applies semantic analysis for identifying ball trajectories and game events by using the domain knowledge such as the expected trajectory of the ball or the change of its direction and speed.

A finite state machine can be eventually used to perform the last step of outcome decision, by using semantic data stored in the previous step and by associating to the sequential evolution of the recognized events the score assignment dictated by the tennis rules.

2.3. Data storage

A huge amount of redundant data is available, but only few processed data are stored in a database in a fine-graded fashion: ball positions, events that change ball trajectory (such as impacts with the ground field or the players' racquet), players' positions, and score assignments (the last one resulting from the outcome decision process). A relational database, namely PostgreSQL, has been chosen to record the information in order to exploit the hierarchical structure of tennis domain data. The purpose is to obtain effective storage while enabling subsequent statistical analyses. Structured meta-data storage enables fast access to key match events, and, at the same time, provides a foundation for combining data in more meaningful ways in the future, thus enabling the system to be customized to coaches needs and easily expandable to accommodate ever-evolving requirements.

3. Processing modules

In this Section we describe all the steps that are performed from the low level processing of the raw data coming from the four synchronized cameras up to the final decision process which collects semantic information and assigns a score.

3.1. Low level processing

This module is responsible for ball and players detection. Several steps are needed to identify them:

1. Image acquisition and storage;
2. Robust background estimation and subtraction;
3. Moving region filtering by connectivity and morphological analysis;
4. Ball and Players candidate selection by spatial and temporal filtering.

The chosen sensors produce data in the form of colour filter arrays using the Bayer pattern (Bayer, 1976). Thus, images coming from the camera are stored in their raw form for archiving purposes. No conversion is done, so that data requires fewer resources to be stored. Colour conversion is applied only on demand for display purposes, while for the subsequent processing raw images are used.

A suitable algorithm was chosen to provide a background estimation and hence extract ball and players candidates with a high confidence level. In particular, the GIVEBACK algorithm is applied as described in Reno et al. (2015). After the background model is learned, the fine tuning procedure continues with a selective mask update, keeping trace of robust foreground areas in which the background is not updated. This enables the algorithm to easily filter ghosts or subjects that stand still on the scene and to retain good player silhouettes. Additional information on the background subtraction approach are available in Reno et al. (2015).

After the background modelling and subtraction, a connected component analysis is applied to the resulting moving regions, to allow an initial estimation of the area dimensions. Then, morphological operations such as dilatation and erosion are used to remove holes and merge neighbour regions. These preliminary steps allow an initial filtering of noisy regions.

Ball candidates are chosen from blobs that have compatible dimensions with the tennis ball. In particular the radius of the inscribed and circumscribed circles of each region is estimated, selecting those having an inner radius between 5 and 10 pixels and an outer radius between 5 and 30 pixels. This is necessary to allow the detection of balls even at high speed, that, due to motion blur, present an ellipsoidal shape. Correct corresponding balls between consecutive frames are found by evaluating the shape features extracted from the two images. For each candidate in one frame, the most similar are first chosen from the other frame. Then, the selection is refined by associating the closest balls in space. Moreover, a speed threshold is applied to filter ball candidates that are not likely to be part of the game. Whenever a match is played, only the fastest ball candidate is chosen, but during training sessions many balls can be thrown simultaneously and processed accordingly.

The selection of player candidates is done with a different set of operations. Since players move at a lower speed and sometimes are in "idle" state, for example waiting for the serve, there is a chance that parts of the silhouette might be considered as background. Even if GIVEBACK algorithm (Reno et al., 2015) reduces the probability of observing such phenomenon, a morphological closing operation can be necessary to consider only big foreground areas as players candidates. It is worth noticing that the chosen background algorithm is able to deal with such situations, as it was developed specifically for this task.

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

Fig. 4. Graphical output of the low level processing in terms of entities coordinates and player silhouette. On the left side, two different symbols are used to mark the entities: X for the player and + for the ball. On the right side, the silhouette of the player is highlighted in red and ball in green. Data from this stage is essential to reconstruct three dimensional coordinates and proceed with the analysis of the match performing the subsequent tasks. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

An example of low level processing result is reported in Fig. 4. On the left image different marks are used to sign the player and the ball, while on the right image the segmented player silhouette is shown in red and the ball in green.

3.2. 3D reconstruction

This module is essentially responsible for providing 3D information of ball and players candidates. Each pair of synchronized images is exploited to produce a sparse point cloud that embeds information about the active entities on the court during the game. The algorithm is mainly composed of the following steps:

1. Homography computation
2. Entity projection on the ground plane
3. 3D information retrieval

First of all, for each pair of cameras observing the opposite half of the field, the homography matrices which map the transformation between the image planes and the ground plane are estimated. A set of reference points placed on the ground plane is measured by a theodolite sensor in a global reference system, and their correspondences in the two image planes are annotated accordingly. Let (X, Y, Z) with $Z = 0$ be the coordinate of the a reference point in the world reference system and (u, v) its corresponding coordinate on the image plane. The general transformation is given in the following equation:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} sX \\ sY \\ s \end{bmatrix} \tag{1}$$

that can be expressed in Cartesian coordinates as:

$$u = \frac{h_{11}X + h_{12}Y + h_{13}}{h_{31}X + h_{32}Y + h_{33}}$$
$$v = \frac{h_{21}X + h_{22}Y + h_{23}}{h_{31}X + h_{32}Y + h_{33}} \tag{2}$$

To estimate the coefficients $h_{i,j}$ at least four corresponding points are needed thus solving the resulting equation system in the least squares sense. Additionally, the theodolite sensor is used to measure the position in the world reference system of the centres of projection (*CP*) of all the cameras. It is worth observing that this procedure is necessary only when installing the system to generate the four homography matrices for the four cameras. Given a point observed in the image plane it is possible to detect the corresponding position *P* on the ground plane and construct the viewing lines between *CP* and *P* as shown in Fig. 5. Whenever the same point is observed by two cameras simultaneously, the intersection of the two viewing lines will then give a 3D point that represents its position in the world reference system. In real cases, each pair of 3D lines constructed this way will not perfectly intersect in a specific point because of noise or numerical approximations. For this reason, the segment of minimum distance between the two skew lines is computed and the 3D point is finally associated to its midpoint.

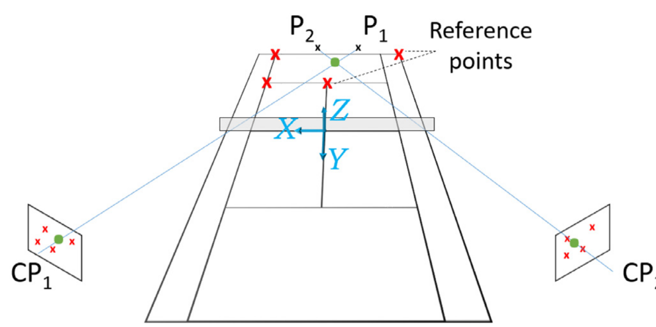

Fig. 5. Example of 3D information retrieval from a pair of homologous cameras. The red points represent some of the reference points chosen on the ground plane which are used to estimate the homography matrices. The green dots represent the ball whose position is determined by the intersection of the two viewing lines (depicted in blue). The global reference system is also shown at the centre of the court, on the ground plane. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3. High level processing

Once 3D positions of ball and players have been retrieved, this module is responsible to segment the acquired sequence into actions. This is initially done by detecting idle parts during the game, where no ball is moving, and splitting the whole sequence accordingly. These chunks of sequences might contain valid game actions or just show a suspended game where inactive balls are collected from the tennis court or exchanged between players while preparing for the next round of serves, a situation very frequent during training sessions. This classification is done after the ball trajectories evaluation.

The temporal analysis of the 3D coordinates of all ball candidates allows the reconstruction of the most complete trajectories which respect the kinematic constraints of the tennis ball. An interpolation procedure is used to approximate the real 3D measures and fill the gap due to missing data and filter out false measures due to noise.

Trajectory breaking is instead concerned with the identification of events affecting the ball. They can be classified as collision with the ground or collision with a player's racquet. These cases modify the ball trajectories and ultimately are events that start a new course for the ball. They are also the key events that are considered by game rules for assigning a score to the action. They can also be easily classified on the basis of their position on the court and of the principal axis where the change of direction is detected. In particular we have that:

- Ball *bounces* are on the ground, and they can be detected by an inversion of speed on the Z axis.
- *Strokes* change the trajectory as well. They affect the ball in all directions and a player must be nearby.
- *Serves* happen on the border of the court when both players assume a specific position and are characterized by the initial height of the trajectory and the high speed.

Particular care in the detection of these trajectory changes must be taken so that noise due to errors of the background subtraction module or to wrong association in the triangulation procedure do not add false positives events. A Gaussian smoothing operation (whose parameters have been experimentally set) is applied before breaking trajectories.

Fig. 6 reports an example of the 3D ball trajectory enriched by separate plots for the (*X, Y, Z*) coordinates. The plot is referred to a complete action that starts with a serve and ends with a scored point, thus the evolution over time of a multiple rally is shown. As reported in Fig. 5, the reference system has the origin in the middle of the court on the ground plane. Trajectory changes along Y are related to *strokes* and local minima and maxima, determined by changes in the sign of the acceleration, can be exploited to understand when a stroke happened. It is worth noting that the Gaussian smoothing operation is useful when dealing with real data, as it reduces the effects due to noise. *Bounces* on the ground are recognizable searching for local minima around 0 in the Z coordinates. *Serves* can be recognized by evaluating players positions on the court along with the height of the ball at the beginning of a tracked trajectory and its Y coordinate (that should be behind the side line), as reported in the box in the figure. Different ball colours in the 3D map of Fig. 6 represent the trajectory over time: blue (serve), cyan, green, yellow, orange and finally red (point). Finally, the dots between the net and the ground (also zoomed in the figure) represent the observations of the ball at the end of the rally.

The resulting trajectory classification is stored in the database together with the estimated frame number and the ball coordinates in the court reference system. At the same time, in correspondence of these events players positions are checked and saved in the database. All this data is used by the following decision process that assigns a score, but can be used for further applications such as query of specific key events, statistics and so on. Trajectories are considered as not valid when there are no strokes in a relatively short period of time, as in Fig. 7 that describes the trajectory of a simple ball pass event between two valid points. Moreover, it is worth noting that the hardware setup described in Section 2 can also introduce time duplicated trajectories when the same ball is captured by both camera pairs (specifically above the net, where all the fields of view are overlapped).

3.4. Outcome decision processing

A score assignment can start only when a complete action has been identified. A complete action is a sequence of frames which starts and ends with idle phases (a number of consecutive frames where no movement of the ball is perceived). The first step searches for a “serve” event, recognized as the first stroke after an

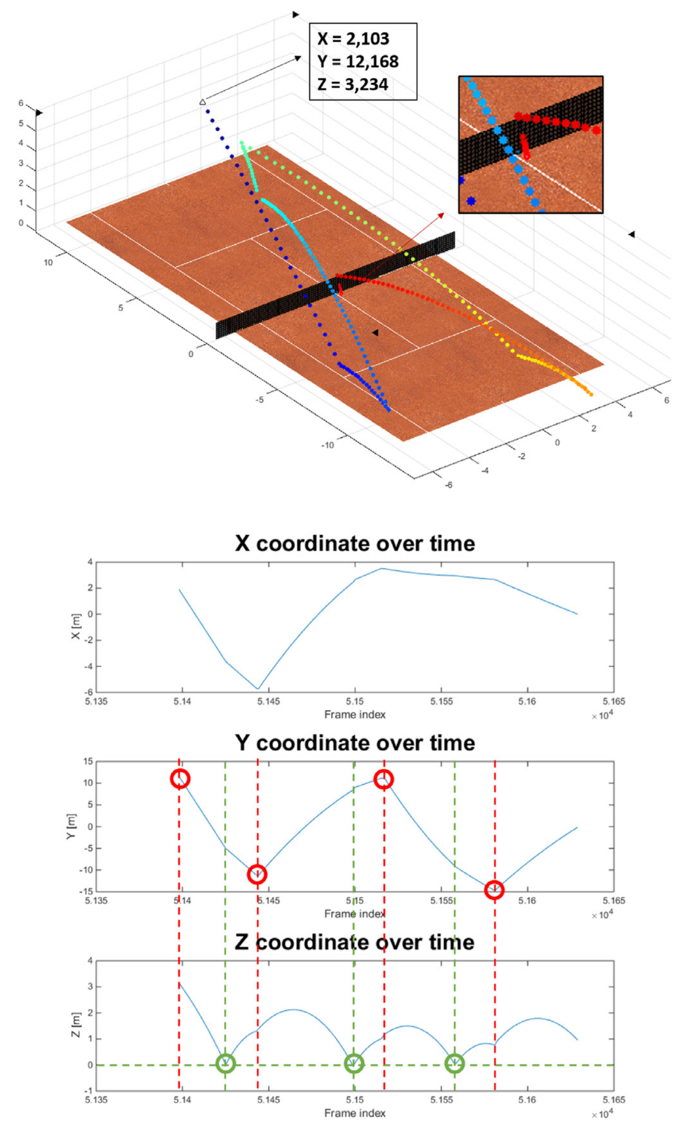

Fig. 6. Example of the 3D ball trajectory and its corresponding X,Y,Z plots studied separately, called action plot. Bounces can be easily recognized looking at the third subplot and searching for local minima around zero (reported as green circles). Strokes can be found exploiting the changes in the sign of the acceleration along the Y-axis (reported as red circles). The box shows the coordinates of the ball in correspondence of the serve. The observations of the ball at the end of the rally are reported in the zoomed rectangle. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

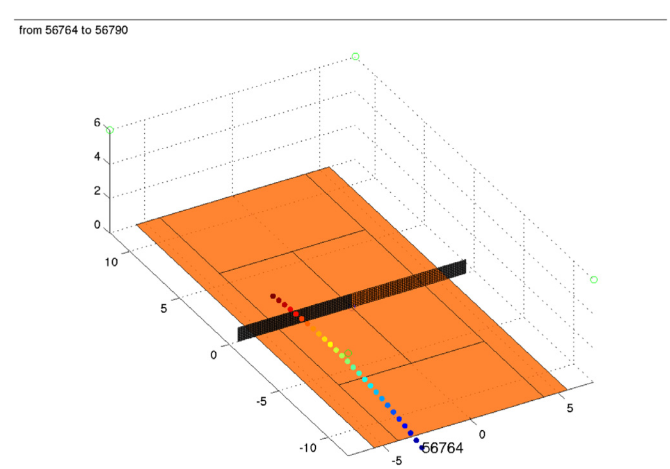

Fig. 7. Example of invalid action: a ball pass between two valid points.

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

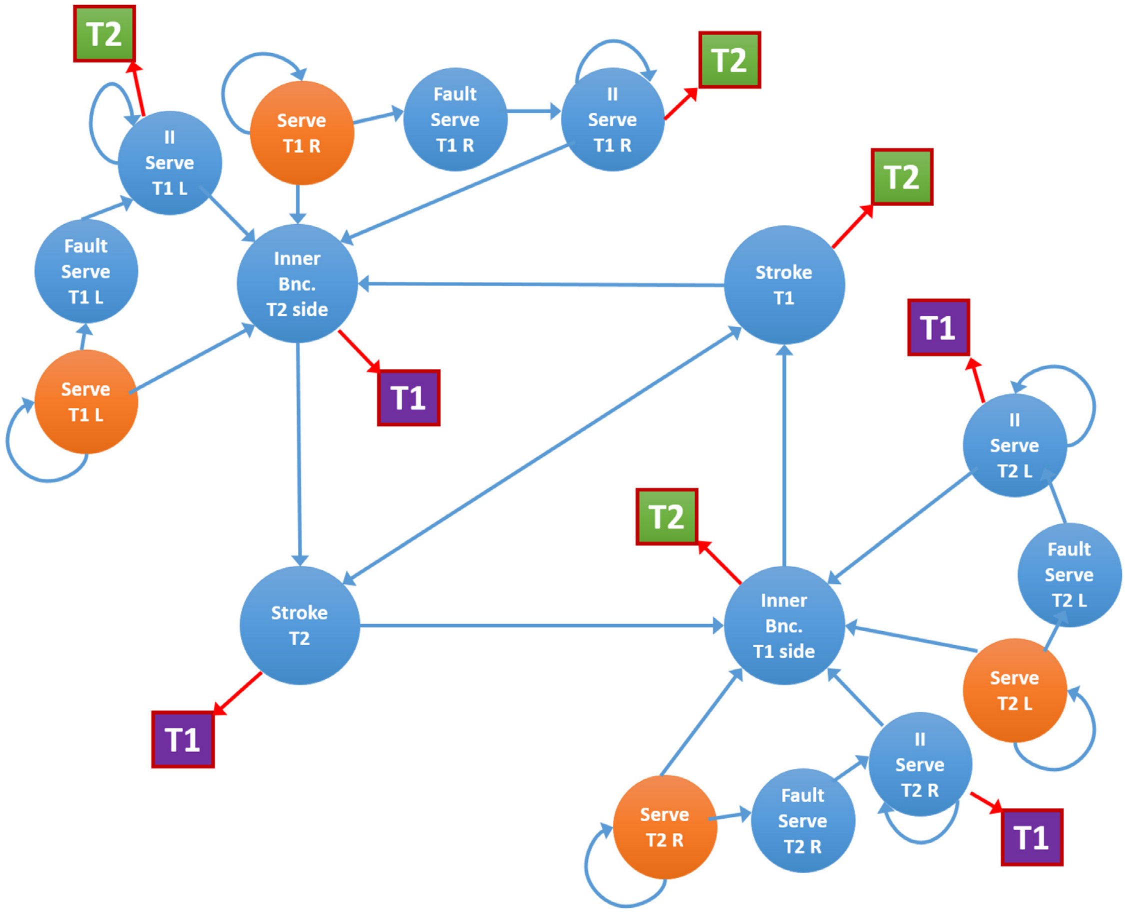

Fig. 8. Graphical model of the finite state machine used to assign a point at the end of each action. The orange filled nodes represent the four initial states that can be found during a tennis match (the serve), while the blue ones are the inner states that describe the progress of the action. The connection between the nodes represent the allowed transitions only. The outcome is reported in squared boxes (purple for team T1 and green for team T2). Each proper event must fire a transition, i.e. the FSM must change its state at each iteration if the action is not concluded. Otherwise, the point is assigned to the green/purple highlighted team in correspondence of the node that did not changed its state. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

idle period that happens near the side line. Then, the ball trajectory is analyzed until the end of the action, when a point is assigned to one of the players. A finite state machine (FSM), which embeds the rules of the game, has been designed. The finite state machine changes the state if the ball follows a valid trajectory with respect to the rules. When the FSM can not reach another valid state in response to an event, the action is considered completed and a point is assigned. Particular attention should be given to the repetition of a serve (first or second) that is allowed only when the served ball touches the net and bounces inside a valid area of the court. In that case, the particular service should not count and the service needs to be repeated without cancelling any previous fault. It should be noted that net events are important in this context only, otherwise they can safely be ignored to correctly assign a score. Fig. 8 shows a graphical overview of the FSM, which resumes all the possible situations that can assign a score, starting from simple aces (for example Serve T1 L, Inner Bounce T2 side → score T1) to more complex actions with several strokes and bounces. In Table 1 all the possible states with their correspondent outcome are reported. The states are extracted by the events stored in the database in the previous step by analyzing both the type of events and the corresponding 3D ball coordinates. It is worth noting that valid court boundaries depend on both game type (single/double) and stroke type (serve or other strokes), therefore the meaning of “inside” and “outside” changes according to the rules of the game.

Table 1
List of all possible FSM states with the correspondent outcome. The * means that no point can be assigned in the specific state. This is true only when a fault occurs and the serve can be repeated.

State	Possible outcome
Serve T1 L/R	*
Fault Serve T1	*
II Serve T1 L/R	T2
Serve T2 L/R	*
Fault Serve T2	*
II Serve T2 L/R	T1
Inner Bounce T1 side	T2
Inner Bounce T2 side	T1
Stroke T1	T2
Stroke T2	T1

In Fig. 8 square boxes represent the key-value map that associates an outcome to the action: if the FSM is not able to change its state, then the point is assigned to the appropriate team.

In order to explain the decision process by the FSM two examples are analyzed. Fig. 9 shows two actions, represented by the sets of events $A_1 = [ev_1, ev_2, ev_3]$ and $A_2 = [ev_1, ev_2, ev_3, ev_4, ev_5]$. Blue lines represent ball trajectories between valid states, while the red lines depict the last state transition in which the decision about the point assignment is made. In the first case the events

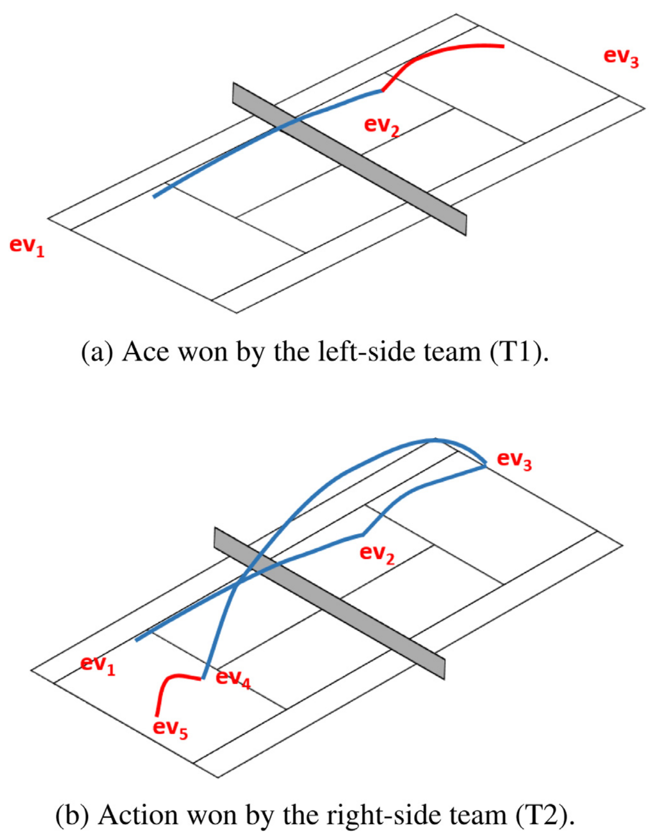

Fig. 9. Examples of actions that can be processed with the FSM to decide the outcome. In the first case, the player on the left side performs an ace and wins the point, while in the second case a longer action is depicted.

Table 2
Analysis of the example action in Fig. 9b. The first event initializes the FSM with a Serve played by team T1 from the right side. Then, other events are received and the machine changes its state according to the representation in Fig. 8. The last state is "Inner bounce T1 side", meaning that the point should be assigned to the team T2 according to the key-map representation in Table 1. The last event is not able to fire a transition, implying the end of the action.

Initial state	**Event**	**Arrival state**
*	(ev_1)	Serve T1 R
Serve T1 R	(ev_2)	Inner bounce T2 side
Inner bounce T2 side	(ev_3)	Stroke T2
Stroke T2	(ev_4)	Inner bounce T1 side
Inner bounce T1 side	(ev_5)	*No valid transition, end of action*

are: serve, bounce and another bounce. The associated states are respectively: Serve T1 R, Inner Bounce T2 side and Inner Bounce T2 side again. The latter state is due to the fact that the second bounce takes place on the same side of the court, therefore since there is no valid state transition, the FSM remains in the previous state. As shown in Fig. 8 the point is assigned to T1.

Following the same approach, the events of the second action are: serve, bounce, stroke, bounce, bounce. The corresponding states are detailed in Table 2. The first stroke event ev_1 is a Serve made from team T1 on the right side of the court, so the initial state is "Serve T1 R". Then, the following events are responsible for allowed state transitions as specified in the Table 2, until the last state Inner bounce T1 side is repeated. Also in this case the last event is another bounce on the same side of the court. Then there is no valid state transition and the point is assigned to T2.

It should be noted that the entrance state of the FSM cannot be necessarily a service, as the proposed system can be used also during training sessions with any kind of action. The FSM in this cases is used to label the observed actions and to store in the data base all the information about number of strokes, positions, bounces, with the resulting scores. These data can be used by coaches to perform useful queries and evaluate player performances both during training and official matches.

Table 3
Experiment 1: the table reports the results in terms of TP, FP, and FN for the recognition of strokes, bounces, and service during a training session of 75 min. In the first column the numbers of events manually labelled during the training sessions. The values of TP, FP, FN, P, R are all percentages.

	GT	TP	FP	FN	P	R
Serve	112	85.8	0.0	14.2	100	85.8
Shot	409	89.6	5.6	4.8	94.1	94.9
Bounce	467	85.9	9.2	4.9	90.3	94.6

4. Experiments and results

The experiments have been performed on a clay court hosted by a private tennis club that has been equipped with the hardware described in Section 2.

Two different experiments have been conducted: the first one to demonstrate the performances of the proposed approach to recognize serves, strokes and bounces, by the analysis of the 3D ball trajectory variations; the second one to test the whole chain, with the FSM that assigns a final score to each action.

In the first case, a number of 225,650 frames have been recorded by four synchronized cameras during different training sessions for a total of about 75 min. In these sessions, players are free to move on the court without respecting strictly rules and under the supervision of their coach, they can improve the technique on particular/unusual strokes, control the length of an action or repeat particular sequences of strokes in order to enhance tactics. The registration of these video sequences has been used to test the ability of the proposed system to segment subsequences which contain actions, track the ball and reconstruct the 3D trajectories in order to recognize services, bounces and strokes.

Before proceeding in the analysis of the results, it should be observed that in this kind of real experiments is actually difficult to establish the ground truth for 3D ball trajectory evaluation. It would be necessary to employ external measurement sensors able to measure exactly the ball trajectory. What it is generally done is to manually label the ball position in the images and reconstruct the 3D ball coordinates by triangulation. In this case also the resulting measures are affected by errors due to the optical projections of the system, the matrices approximations, and so on. The

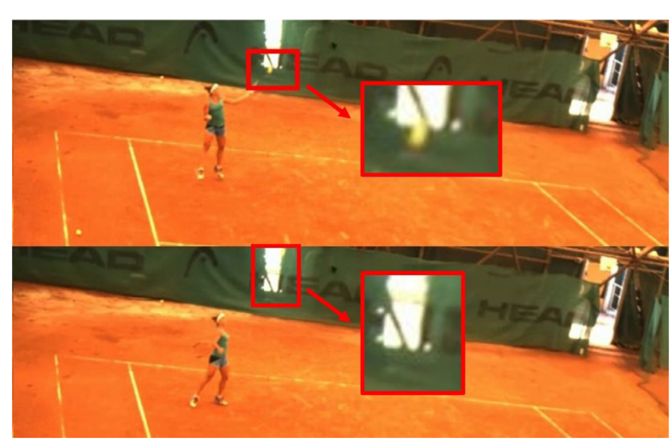

Fig. 10. Examples of images in which the ball is not visible by one of the cameras.

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

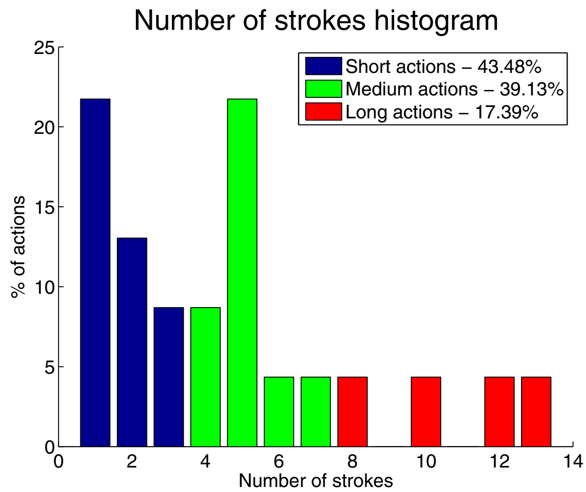

Fig. 11. Number of strokes histogram.

scope of this paper is not to evaluate exactly the 3D ball trajectory, but only to establish the ability of the system to recognize events. The ground truth has been then generated by manually labelling the frames in which users recognize strokes, bounces and services, and comparing these results with those produced by the system. The proposed system was tested on very long sequences, but results are reported only on 75 min of recording where the manually generated ground truth was available.

In Table 3 the results in terms of True Positives, False Positives, False Negatives, Precision and Recall are provided for a total of 112 services, 409 strokes, and 467 bounces. We have considered as correct detection when an event is recognized within a temporal window of 20 frames corresponding to 0.5 s which seems a reasonable interval in which different events cannot be found. The detection in terms of true positives are satisfactory for all the kinds of events. In particular, the number of services which are not recognized are due to the fact that the ball tracking starts later than the actual ones, and the constraints for the service detection are not met anymore (a service is a stroke that happens near the

Table 4
This table shows the details about actions, one per each row. Validity and events count are reported as well as temporal information in terms of starting frame, ending frame and duration, compared with manually annotated ground truth. Invalid actions are characterized by one of the following: absence of strokes in combination with a relatively short duration, sub actions whose temporal boundaries are intersected with a main action (as explained in Section 4) or simple ball pass events between two points. Finally, three missing serves have been highlighted in bold.

	Ground truth					System output							
#	Start frame	End frame	Srv	Str	Bnc	Start frame	End frame	Srv	Str	Bnc	Delta start	Delta end	Decision
01	47,855	47,964	1	1	1	47,852	47,916	1	1	1	3	48	Valid
02	48,387	48,559	1	2	2	48,380	48,592	1	2	2	7	−33	Valid
03	49,524	49,693	1	2	2	49,519	49,651	1	2	2	5	42	Valid
04	50,569	50,918	1	3	4	50,589	50,875	0	3	4	−20	329	**Missing serve**
05	51,402	51,658	1	3	3	51,393	51,629	1	3	3	9	29	Valid
06	–	–	0	0	2	52,015	52,211	0	0	2	–	–	Ball Pass
07	52,648	52,840	1	4	5	52,667	52,984	0	4	5	−19	−144	**Missing serve**
08	54,751	55,722	1	12	12	54,747	55,763	1	12	12	4	−41	Valid
09	–	–	0	0	0	54,879	54,912	0	0	0	–	–	Overlap
10	56,425	57,129	1	9	9	56,416	57,135	1	9	9	9	−6	Valid
11	–	–	0	0	0	56,764	56,790	0	0	0	–	–	Overlap
12	58,133	58,178	1	0	1	58,126	58,178	1	0	1	7	0	Valid
13	–	–	0	1	1	58,424	58,546	0	1	1	–	–	Ball pass
14	59,002	59,360	1	4	5	58,945	59,323	1	4	5	57	37	Valid
15	–	–	0	0	0	59,181	59,223	0	0	0	–	–	Overlap
16	–	–	0	1	2	59,546	59,758	0	1	2	–	–	Ball pass
17	60,731	61,350	1	4	5	60,673	61,099	1	4	5	58	251	Valid
18	–	–	0	1	1	61,095	61,333	0	1	1	–	–	Overlap
19	–	–	0	0	1	61,790	61,925	0	0	1	–	–	Ball pass
20	–	–	0	1	1	61,677	61,791	0	1	1	–	–	Ball pass
21	–	–	0	0	0	62,348	62,483	0	0	0	–	–	Ball pass
22	63,615	64,008	1	5	4	63,612	64,001	1	5	4	3	7	Valid
23	–	–	0	1	1	64,521	64,664	0	1	1	–	–	Ball pass
24	–	–	0	0	1	64,633	64,782	0	0	1	–	–	Ball pass
25	65,213	65,561	1	4	5	65,206	65,562	1	4	5	7	−1	Valid
26	–	–	0	1	3	65,892	66,106	0	1	3	–	–	Ball pass
27	–	–	0	0	0	65,944	66,065	0	0	0	–	–	Ball pass
28	66,525	66,567	1	0	1	66,521	66,553	1	0	1	4	14	Valid
29	66,968	67,189	1	0	1	66,970	67,002	1	0	1	−2	187	Valid
30	–	–	0	1	2	67,612	67,758	0	1	2	–	–	Ball pass
31	68,113	68,956	1	11	11	68,107	68,928	1	11	11	6	28	Valid
32	–	–	0	0	0	68,581	68,609	0	0	0	–	–	Overlap
33	70,092	70,197	1	1	1	70,088	70,171	1	1	1	4	26	Valid
34	70,891	71,003	1	1	2	70,895	71,017	1	1	2	−4	−14	Valid
35	72,449	72,509	1	0	2	72,443	72,499	1	0	2	6	10	Valid
36	72,705	73,241	1	7	7	72,702	73,228	1	7	7	3	13	Valid
37	74,776	75,287	1	6	7	74,711	75,271	1	6	7	65	16	Valid
38	–	–	0	0	0	75,055	75,087	0	0	0	–	–	Overlap
39	–	–	0	1	1	75,551	75,738	0	1	1	–	–	Ball pass
40	76,289	76,357	1	0	1	76,227	76,335	1	0	1	62	22	Valid
41	–	–	0	0	2	76,447	76,604	0	0	2	–	–	Ball pass
42	78,415	78,835	1	4	5	78,444	78,747	0	4	5	−29	88	**Missing serve**
43	–	–	0	1	1	84,498	84,627	0	1	1	–	–	Match end
44	–	–	0	0	2	85,286	85,469	0	0	2	–	–	Match end
45	–	–	0	2	1	85,821	85,933	0	2	1	–	–	Match end

Table 5
The table reports the action number, the initial state of the FSM, the outcome of each action (Win) assigned by the FSM, and the score of the actions that could be assigned automatically. The notation W is used to identify the game point.

Action	In. State	Win	Game	T1	T2
01	Serve T2 R	Fault	1	0	0
02	II Serve T2 R	T2	1	0	15
03	Serve T2 L	T1	1	15	15
04	In Bnc T1	T2	1	15	30
05	Serve T2 L	T2	1	15	40
07	In Bnc T1	T2	1	15	W
08	Serve T2 R	T1	2	15	0
10	Serve T2 L	T2	2	15	15
12	Serve T2 R	T1	2	30	15
14	Serve T2 L	T2	2	30	30
17	Serve T2 R	T1	2	40	30
22	Serve T2 L	T1	2	W	30
25	Serve T1 R	T2	3	0	15
28	Serve T1 L	Fault	3	0	15
29	II Serve T1 L	T1	3	15	15
31	Serve T1 R	T2	3	15	30
33	Serve T1 L	T2	3	15	40
34	Serve T1 R	T2	3	15	W
35	Serve T1 R	Fault	4	0	0
36	II Serve T1 R	T1	4	15	0
37	Serve T1 L	T1	4	30	0
40	Serve T1 R	T1	4	40	0
42	In Bnc T2	T1	4	W	0

bottom line of the tennis court and with a certain elevation with respect to the ground). In the shot detection there is a percentage of False Positive due to the fact that the analysis of the changes in the sign of acceleration can be caused by local minima which were not filtered by the smoothing step. Also for the bounce detection, some false positives are due to strokes which happen very close to the field and the constraint on the minimum around 0 in the Z coordinate fails. False negatives are generally due to the not precise trajectory reconstruction due to failures in the ball detection in some frames. For example in Fig. 10 one of the cameras cannot see the ball as it is saturated by the lights on the advertisement. The system overall performance can be then evaluated in terms of Precision and Recall percentages that have been obtained during the experiment. The first value indicates how many selected items are relevant, while the second one expresses how many relevant items have been output by the system. For the categories Shot and Bounce the system is able to achieve values greater than 90%, proving its capability to output a high number of True Positive values along with a low number of False Positive or Negatives. The Serve Recall value scores 85.8% and is degraded by the False Negative values that have been discussed before.

Another point which should be considered for the evaluation of the system is the precision in the evaluation of the ball position when bounces are recognized as these data are necessary to assess if the ball is outside or inside the valid court. Certainly the performances of the proposed systems cannot be compared to complex commercial ones (such as Hawkeye (Owens et al., 2003)) based on a greater number of cameras which observe only the lines and perform 3D trajectory reconstruction. Anyway in order to have a general idea about these measurement errors, we have manually labelled the position of the ball in correspondence of the observed bounces, and compared the ball position obtained by our system with respect to those estimated by the manually annotation procedure. The same observation made beforehand about the ground truth is still valid. Also these measures are affected by noise, as the variation of one pixel in the ball manual labelling produces variations in the ball localization of several centimetres. For this reason, comparative results can be considered only in a qualitative way. Indeed, for the 91% of bounces the ball position error is estimated

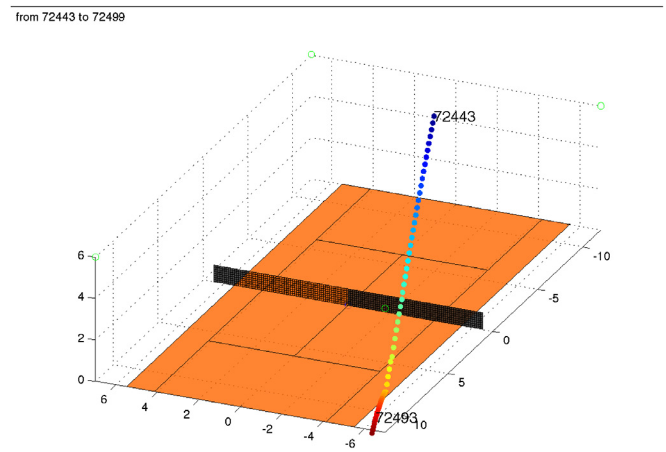

Fig. 12. Example of wrong serve in which the bounce is outside the allowed area.

under 15 cm. As a consequence, when the ball is close to border of the valid field, the system could fail in determining inside or outside situations.

At this point the actions can be preliminarily analyzed for statistical purposes grouping them with respect to the number of strokes that occur during the play of a single point. Fig. 11 shows the distribution of the actions according to this representation:

- short duration ones are reported in blue and cover about 43% of the total number of actions and generally refer to faults, aces or points that finish just after a couple of strokes;
- medium duration actions are the green ones, represent about 39% and are related to well balanced points in which both players are playing similarly;
- long duration actions, the red ones, are only 17% and can be exploited similarly to the previous ones.

This kind of statistics can be used by coaches to evaluate performances and extract video sequences containing specific events such as actions with one stroke (probably corresponding to fault or winning return), actions with long exchanges, and so on.

The second experiment consisted of 38, 000 synchronized frames that cover a real match made of four games. In this experiment the finite state machine has been tested to assign points and keep track of the score. It is important to put in evidence that the FSM just embeds information on game rules that are fixed. Therefore, its behaviour is deterministic. For this reason, the purpose of this second part of the experiments is to understand whether this system will be able to effectively assist a coach with a high confidence level.

A total number of 45 actions automatically tagged as valid or invalid by the system have been identified.

Table 4 contains the details about each action. We report the manually labelled ground truth (Start frame, End frame, Srv, Str, Bnc), the corresponding system output, and the final evaluations carried out by our system. The ground truth in terms of start frame and end frame has been reported only for the valid actions, while for ball pass actions or idle periods are not reported. Strokes and bounces are correctly detected, while in three cases serves are not recognized (actions 4,7, and 42). Ground truth start frame is correctly identified within a certain number of frames, showing that the system can effectively segment actions. Only for the three actions in which serves were not recognized, the starting frames were found later than the actual ones. This is a clear indication that the ball was not visible initially and the tracked trajectory started later causing a failure of the constraints for the serve detection. Moreover, the 1: 1 correspondence between the evaluated

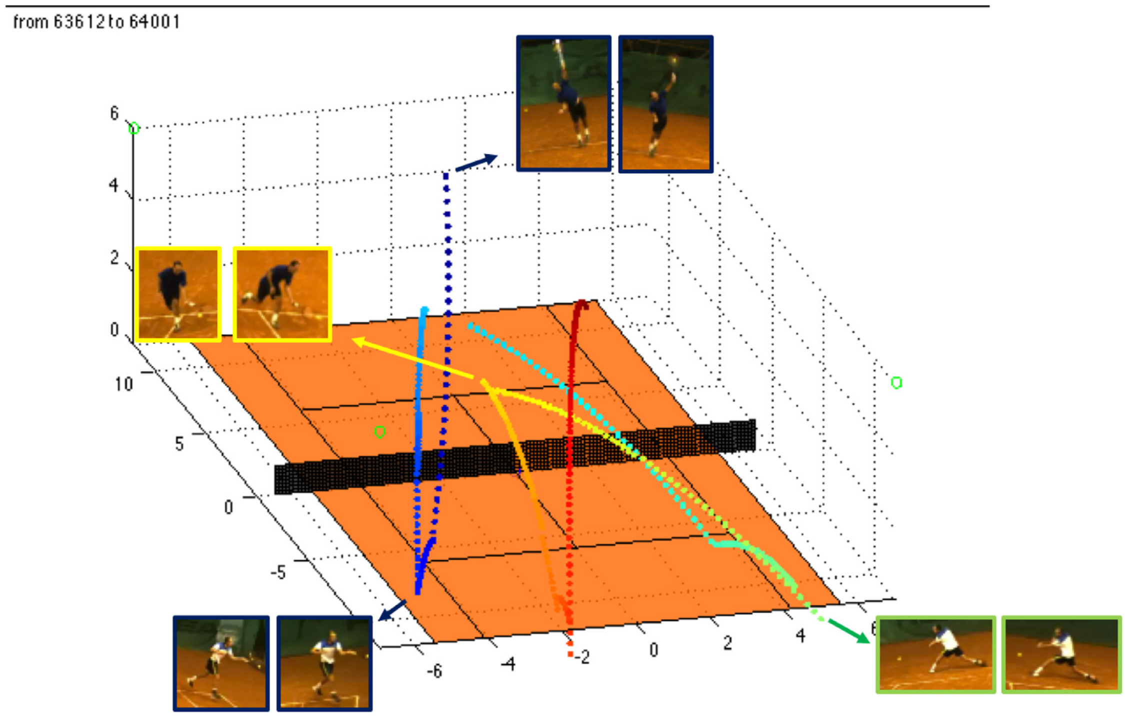

Fig. 13. Example of full action with two relevant strokes highlighted per each player. Since the database contains an event-indexed representation of the match, it is extremely simple to seek and view key events during a match.

outcome and ground truth data confirms the capability of identifying bounces and strokes correctly. Action number 27 is an example of short trajectory fully contained in action 26 that is neglected because it is a duplicate one. Actions are considered valid if they start with a serve. When a serve is not recognized and actions have a long duration with a high number of strokes and bounces, actions are considered still valid but labelled with Missing serve.

Then, the 23 valid actions (20 valid and 3 missing serves) extracted by the previous high level processing, have been analyzed in order to automatically annotate the score. Table 5 shows the action number, the initial state of the FSM, the outcome of each action (Win) assigned by the FSM, and the score of the games that could be assigned automatically. Although in the actions 4,7, and 42 the initial serves were missed and the initial states of the FSM were inner bounces T1 or T2 , the scores were correctly assigned to the correct team. In this case, as the initial preprocessing correctly recognizes the other events, the score assignment in the considered sequence of games is performed correctly as well.

Some examples of the reconstructed actions with the ball trajectories plotted in a 3D Euclidean space are reported in Figs. 12 and 13. Fig. 12 represents a fault where the ball bounces is outside the side line. Fig. 13 represents action number 22. The colours change according to the frame index (blue, cyan, green, yellow, orange and finally red). Some relevant strokes are highlighted in the Fig. 13 and represented as players are seen by the respective cameras. The serve is the starting event of the action and is depicted in blue: the player seen in blue assumes the typical serve position (similar to a smash) behind the side line. Other three examples of strokes are provided in the same Figure: the return, that takes place outside the single sideline; a shot (highlighted in green) played by the white-dressed player on the extreme right side of the court and finally a stroke between the service line and the net (yellow rectangle). The miniatures of players demonstrate the potential use of the data: coaches can perform intelligent queries to the database, can extract specific actions and analyze just the frames in which players hit the ball.

The considered sequence of games was used to demonstrate that the FSM is able to decide correctly the score assignment whenever the preliminary decision process (event recognition) is correct. Anyway, as discussed above, in some cases bounces can be confused or assigned erroneously inside/outside. In these cases the score could be wrong. However, since the scope of the paper is neither to do extremely precise measures nor automatic score assignment but to allow coaches to extract interesting video sequences for player performances analysis, it is not relevant if some actions terminate with a wrong score assignment. The wrong decision can be manually adjusted by the coach in a second time. As the FSM can be used also during training sessions to label extracted actions with scores, coaches can save long time analyzing only short video sequences which contain significant exchanges instead of observing all the recorded sequences. In particular, coaches can exploit this system functionality to filter relevant parts of the recorded sequences according to their training strategies. As illustrative examples, to improve the attacking capabilities of players, all the events that occur in the attack zone can be selected, or to evaluate the reacting capabilities, consecutive actions with lost scores can be analyzed.

Software modules have been developed in C++ and Matlab languages. In particular, low level processing is the most computationally expensive task, as it must run on each raw video frame. The current implementation of this module runs at 30 fps and will certainly benefit from further optimizations. Once 3D information are extracted from the low level processing module, the high level processing (trajectories processing, events recognition and outcome decision) are performed in Matlab environment and do not require further optimization. However, the whole system architecture will likely benefit from the integration of all these modules in the same language.

5. Conclusion and future works

In this paper we propose a visual system based on four synchronized cameras which is able to record training and official tennis matches, segment action in frame sequences, recognize significant events such as strokes, bounces or services, and eventually assign a final score. The system has been designed to meet requirements coming from domain experts. It can be used by coaches and players to analyze long training sessions, and without observing all the sequences, extract significant actions, such as those ending with a positive score or containing at least a certain number of strokes, etc. up to our knowledge, it is the first system which tries to segment video sequences while adding semantic information useful for player performance analysis.

The system integration phase involved an accurate hardware choice for making the proposed solution modular, scalable and flexible at the same time. Design and implementation of software integrated solutions have been investigated as well, in order to obtain an event based indexed representation of a match starting from big raw data acquired from cameras. A remarkable feature of the whole approach consists in the absence of invasiveness: players are simply free to behave like they already do while the system does acquisition and processing differently from wearable-based solutions.

Since not all the software modules effectively can reach and maintain real time performances, a future improvement of the whole pipeline will regard optimized hardware implementations (e.g. using FPGA cards or GPU arrays) to speed up computations. Future works will also be devoted to the identification of the type of the strokes exploiting 3D information about both ball and players. For example, it will be useful to label a stroke as forehand or backhand, but also lob, drop shot, smash or volley. Such high level data will be exploited to enrich the analysis of both game tactics and players intentions.

Acknowledgements

The authors would like to thank the anonymous reviewers whose valuable comments contributed to the critical improvement of the paper and Mr. Michele Attolico for his technical support during real experiments.

References

Ahmadi, A., Rowlands, D., James, D.A., 2009. Towards a wearable device for skill assessment and skill acquisition of a tennis player during the first serve. Sports Technol. 2 (3–4), 129–136.
Archana, M., Geetha, M., 2015. Object detection and tracking based on trajectory in broadcast tennis video. Procedia. Comput. Sci. 58, 225–232.
Avenir sports 2016. URL http://avenirsports.ie/.
Bächlin, M., Förster, K., Tröster, G., 2009. Swimmaster: a wearable assistant for swimmer. In: Proceedings of the 11th International Conference on Ubiquitous Computing. ACM, pp. 215–224.
Bayer, B. E., 1976. Color imaging array. US Patent 3,971,065.
Bloom, T., Bradley, A.P., 2003. Player tracking and stroke recognition in tennis video. In: APRS Workshop on Digital Image Computing (WDIC'03), Vol. 1. The University of Queensland, pp. 93–97.
Chen, C., Pomalaza-Ráez, C., 2009. Monitoring human movements at home using wearable wireless sensors. In: Proceedings of the Third International Symposium on Medical Information and Communication Technology.
Chi, E.H., 2005. Introducing wearable force sensors in martial arts. Pervasive Comput. IEEE 4 (3), 47–53.
Conaire, C.O., Kelly, P., Connaghan, D., O'Connor, N.E., 2009. Tennissense: a platform for extracting semantic information from multi-camera tennis data. In: Digital Signal Processing, 2009 16th International Conference on. IEEE, pp. 1–6.
Connaghan, D., Kelly, P., O'Connor, N.E., 2011. Game, shot and match: event-based indexing of tennis. In: Content-Based Multimedia Indexing (CBMI), 2011 9th International Workshop on. IEEE, pp. 97–102.
Dartfish. 905 URL http://www.dartfish.com/en/index.htm.
D'Orazio, T., Leo, M., 2010. A review of vision-based systems for soccer video analysis. Pattern Recognit. 43 (8), 2911–2926.
D'Orazio, T., Leo, M., Mazzeo, P., Mosca, N., Nitti, M., Distante, A., 2009. An investigation into the feasibility of real-time soccer offside detection from a multiple camera system. IEEE Trans. Circuits Syst. Video Technol. 19 (12), 1804–1818.
D'Orazio, T., Leo, M., Spagnolo, P., Nitti, M., Mosca, N., Distante, A., 2009. A visual system for real time detection of goal events during soccer matches. Comput. Vision Image Understanding 113 (5), 622–632.
Ermes, M., Pärkkä, J., Mäntyjärvi, J., Korhonen, I., 2008. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. IEEE Trans. Inf. Technol. Biomed. 12 (1), 20–26. doi:10.1109/TITB.2007.899496.
Ghasemzadeh, H., Loseu, V., Jafari, R., 2009. Wearable coach for sport training: a quantitative model to evaluate wrist-rotation in golf. J. Ambient Intell. Smart Environ. 1 (2), 173–184.
Hughes, M., Franks, I.M., 2004. Notational Analysis of Sport: Systems for Better Coaching and Performance in Sport. Psychology Press.
Kapela, R., Swietlicka, A., Rybarczyk, A., Kolanowski, K., O'Connor, N., 2015. Real–time event classification in field sport videos. Signal Process. Image Commun. 35, 25–45.
Kokaram, A., Rea, N., Dahyot, R., Tekalp, A.M., Bouthemy, P., Gros, P., Sezan, I., 2006. Browsing sports video: trends in sports-related indexing and retrieval work. Signal Process. Mag. IEEE 23 (2), 47–58. doi:10.1109/MSP.2006.1621448.
Lai, J.H., Chen, C.H., Kao, C.C., Chien, S.Y., 2011. Tennis video 2.0: a new presentation of sports videos with content separation and rendering. J. Vis. Commun. Image R. 22, 271–283.
Leo, M., Mosca, N., Mazzeo, P., Nitti, M., D'Orazio, T., Distante, A., 2008. Real-time multiview analysis of soccer matches for understanding interactions between ball and players. In: Proceedings of the 2008 International Conference on Content-Based Image and Video Retrieval, pp. 525–534.
Match Analysis, 2016. Match analysis. URL http://matchanalysis.com/.
Owens, N., Harris, C., Stennett, C., 2003. Hawk-eye tennis system. In: IEE Conference Publication. Institution of Electrical Engineers, pp. 182–185.
Performa Sports, 2016. Performa sports. URL http://www.performasports.com/.
Pingali, G., Jean, Y., Carlbom, I., 1999. Lucent vision: a system for enhanced sports viewing, volume 1614 of. Lect. Notes Comput. Sci. 689–696.
Pingali, G., Opalach, A., Jean, Y., 2000. Ball tracking and virtual replays for innovative tennis broadcasts. In: Pattern Recognition, 2000. Proceedings. 15th International Conference on, Vol. 4. IEEE, pp. 152–156.
Poliakov, A., Marraud, D., Reithler, L., Chatain, C., 2010. Physics based 3d ball tracking for tennis videos. In: Content-Based Multimedia Indexing (CBMI), 2010 International Workshop on. IEEE, pp. 1–6.
Polk, T., Yang, J., Hu, Y., Zhao, Y., 2014. Tennivis: visualization for tennis match analysis. IEEE Trans. Vis. Comput. Graph. 20 (12), 225–232.
Protracker Tennis, 2015. Protracker tennis. URL http://www.fieldtown.co.uk/.
Reno, V., Mosca, N., Nitti, M., D'Orazio, T., Campagnoli, D., Prati, A., Stella, E., 2015. Tennis player segmentation for semantic behavior analysis. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 1–8.
Strohrmann, C., Harms, H., Tröster, G., Hensler, S., Müller, R., 2011. Out of the lab and into the woods: kinematic analysis in running using wearable sensors. In: Proceedings of the 13th International Conference on Ubiquitous Computing. ACM, pp. 119–122.
Tamaki, S., Saito, H., 2013. Reconstruction of 3d trajectories for performance analysis in table tennis. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on. IEEE, pp. 1019–1026.
Valter, D.S., Adam, C., Barry, M., Marco, C., 2006. Validation of prozone®: a new video-based performance analysis system. Int. J. Perform. Anal. Sport 6 (1), 108–119.
Yan, F., Christmas, W., Kittler, J., 2005. A tennis ball tracking algorithm for automatic annotation of tennis match. In: British Machine Vision Conference, Vol. 2, pp. 619–628.
Yan, F., Kostin, A., Christmas, W., Kittler, J., 2006. A novel data association algorithm for object tracking in clutter with application to tennis video analysis. In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, Vol. 1. IEEE, pp. 634–641.
Yu, X., Xu, C., Leong, H.W., Tian, Q., Tang, Q., Wan, K.W., 2003. Trajectory-based ball detection and tracking with applications to semantic analysis of broadcast soccer video. In: Proceedings of the Eleventh ACM International Conference on Multimedia. ACM, pp. 11–20.