



A Low-Cost Computer Vision System for Real-Time Tennis Analysis

S. Messelodi¹(✉), C. M. Modena¹, V. Ropele², S. Marcon², and M. Sgrò²

¹ FBK-irst, Via Sommarive 18, 38123 Trento, Italy
{messelod,modena}@fbk.eu

² Eyes VisiON srl, Via alla Cascata, 56C, 38123 Trento, Italy
{valerio.ropele,stefano.marcon,mario.sgro}@eyeson.tennis

Abstract. This paper describes a low-cost vision-based system for real-time tennis game analysis. The system elaborates videos captured by four synchronized and calibrated cameras installed at the sides of the court in order to accurately localize ball and players, and track them in real-time. From this low-level data mid-level events, like shots, bounces, ball in net, and high-level events, like stroke type and line calling, are detected. All this data is made available to the players both on-court during the play or through a web device at the end of the session. Currently, system prototypes are undergoing a field test in three locations in Italy. In addition to positive comments of users, robustness and reliability of the system have been demonstrated with specific evaluation tests. Detection rate of shots is 99.7% while miss detection rate is less than 0.8%. Reliability of the stroke classification is 97.1% and of *in/out* evaluation is 99.5%. On average reaction time for line calling is 152 ms.

Keywords: Tennis analysis · Real-time image processing · Line calling

1 Introduction

Sport is a physical activity aiming to improve skills through personal challenges and competitions and, at the same time, to provide recreation and enjoyment to participants. Competitive sport represents a source of entertainment for an ever increasing number of spectators around the world, especially thanks to the wide diffusion of television and digital media, making it an important business.

The topic of content-based analysis of broadcast sport material, including tennis, has received, and continues to receive, great attention and efforts from the computer vision research community. In particular, techniques have been proposed for automatic annotation of multimedia archives for information retrieval [19,21] and for video enhancement [16] using augmented reality methods to improve the engagement of spectators.

This work has been partially funded by Provincia Autonoma di Trento (Italy) under L.P. 6/99. Authors would like to thank the staff involved in the system development.

We propose a low-cost solution based on computer vision to extract in real-time low-level information from a tennis game on-court. Specifically, our system is able to detect and track the 3D position of ball and players. This data is then used by a higher level module to extract analytics than can be offered to the players both on the field and through a web application. Currently the system provides (i) classified detection of shots (forehand, backhand, serve) along with information about space/time localization, velocity and spin, (ii) detection of ball bounces with the estimation of the court contact region supporting the line calling functionality. The solution has been included in a novel system, called EYES ON¹ with the main purpose to allow players to experience the game in a novel and appealing modality, by checking their personal performance, their progress over time, and to get help in the evaluation of controversial in/out cases.

The paper is organized as follows: Sect. 2 provides a brief overview of the state of the art of products or systems devoted to the tennis world. Section 3 presents the main features and the general architecture of EYES ON. Section 4 provides a description of the computer vision sub-system, its modules and the low-level data they provide. Section 5 describes the analytics that are computed from data collected by the vision sub-system, and their visualization through a user interface. Experiments and system performance are provided in Sect. 6, while Sect. 7 concludes the paper with planned future upgrades.

2 Related Work

In the last two decades products devoted to tennis analysis with dedicated cameras in the court have been proposed. For example, LucentVision [15,16], Hawk-Eye [14], and Foxtenn are professional monitoring tennis systems. LucentVision, a system for enhanced tennis broadcasts using real-time game statistics and virtual replays, was launched in 1998 with the ATP Tour and has been used in broadcasts of international tennis tournaments. Hawk-Eye [1,14] ball tracking system is the most advanced tool used in official tennis matches since 2002, and still in use today. It is known for its electronic line calling functionality. Hawk-Eye uses up to ten high resolution cameras placed on the stadium roof, capturing images at 50/60 frame per second (fps) [12,20]. Foxtenn, proposed by Foxtennis Begreen, includes the line calling functionality and recently it received the official approval from the International Tennis Federation (ITF) [11] for its usage in competitions. The system uses up to 8 cameras at 120/2500 fps synchronized with a high-speed laser scanner system. Because of their cost and the complex installation and calibration procedure, they are targeted to the professional circuit. Recently, some monitoring products addressed to players appeared on the market, but aiming only at specific practice or training task, e.g. monitoring the hits of balls being shot from a ball machine [22], supporting coaches for player performance analysis [4,18], detecting where a ball has landed on the court to support the line calling function [9], with a camera mounted on a net-post.

¹ <https://www.eyeson.tennis>.

A few systems for monitoring tennis game have been proposed in the literature. A comprehensive system is TennisSense [3], an instrumented platform for indoor courts, devoted to player performance analysis and health monitoring. The project focuses on multi-modal sensors integration, including a computer vision system and wearable electronic sensing devices. Among the systems based solely on vision we cite the video indexing system outlined in [17] and the platform described in [18]. In both cases they suffer the limitation to operate only for indoor court and, most important, not in real-time.

To the best of our knowledge SmartCourt [10], offered by PlaySight Interactive, is the only system on the market that provides a complete monitoring of tennis game and that can be compared to EYES ON, at least in terms of functionalities. It is equipped with four high-definition cameras working at 50 fps, one or more web cameras, and a graphical interactive interface located in a kiosk structure next to the court. Cameras must be positioned at the top of poles placed at the four corners of the court. It employs image processing algorithms to extract information about strokes, ball trajectory, speed of shots, as well as player movements. Purpose of the web cameras is to support video recording for instant replay on the kiosk from different angles.

3 Outline of the System

The design and implementation of the proposed computer vision system took into account a list of requirements of the whole project: (i) extraction of useful information, part of which immediately available to the players, (ii) accuracy of detection in order to enable the line calling functionality, (iii) applicability in a wide range of different conditions, (iv) ease of installation and configuration, and (v) low-cost. The global architecture of EYES ON is schematized in Fig. 1. The core resides in Vision System whose task is to process video streams provided by cameras observing the court for a twofold purpose: (i) extract low-level data related to 3D position of ball and players with a sufficient temporal resolution to generate accurate trajectories, and (ii) detect relevant events - like ball bounce, ball in net, ball hit by racket - along with their localization in space and time. This data, as soon as it is available, is sent to a Supervisor module which controls the whole system by managing (i) the interaction with the user and (ii) the communication with a local database and the Cloud. The Supervisor includes a video analytics module for the processing of low-level data coming from the Vision System to extract high-level meaningful information for the users, e.g. stroke classification or in/out decision. The interaction between Supervisor and user, through the on-court GUI, allows players to register and to select the desired game modality: match, warm-up, or drills. According to the selected mode the Supervisor properly configures the analytics module. Computed analytics are stored in a local database and, upon request, are shown to the players through the GUI. At the end of each session the extracted data is sent to the Cloud and stored in a permanent database that can be remotely accessed to get personalized performance analysis, trends, or comparison with players community. The hardware configuration of the system includes a processing unit hosted

in a cabinet on one side of the net along with a touch screen display, and four cameras that, in a common set-up, are mounted on two supports located next to the net-posts. The two halves of the court are monitored independently by pairs of cameras. They are synchronized and cable-connected to the processing unit. A position close to the net, besides simplifying the installation, helps in minimizing ball occlusions due to the body of players. But, as a matter of fact, cameras do not necessarily have to be mounted on such supports, and they can be installed on possibly existing structures provided they offer a comparable view of the scene. Other cameras may be optionally installed in the court with the aim to observe and record tennis action from different points of view.

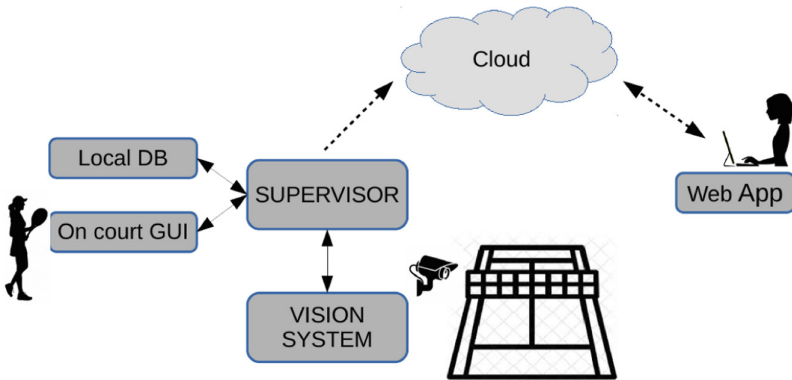


Fig. 1. EYES ON's software architecture. Through the On-court GUI players can register and select the game mode; Supervisor activates the Vision System and, accordingly to game mode, computes and stores analytics on local and cloud DBs. Results are accessible immediately through On-court GUI or via Web App at the end of the session.

4 Vision System

To meet system requirements we designed and implemented computer vision modules characterized by: (i) low complexity in order to run in real-time, (ii) adaptability to a wide range of illumination conditions, (iii) flexibility with respect to moderate variations of camera position, (iv) ability to detect and track accurately small and fast moving objects (e.g. 180 km/h). The Vision System is organized as a set of different specialized software modules which are coordinated at run time by a Vision Manager (VM), which is also in charge of communicating with the Supervisor. The software architecture is diagrammed in Fig. 2. Software modules run on different threads whose execution is orchestrated by VM. VM keeps track of the global state of the monitoring system including information like: a ball is currently tracked or both the players are in the court. In the following we describe different functionalities provided by the Vision System along with the role played by the modules in Fig. 2.

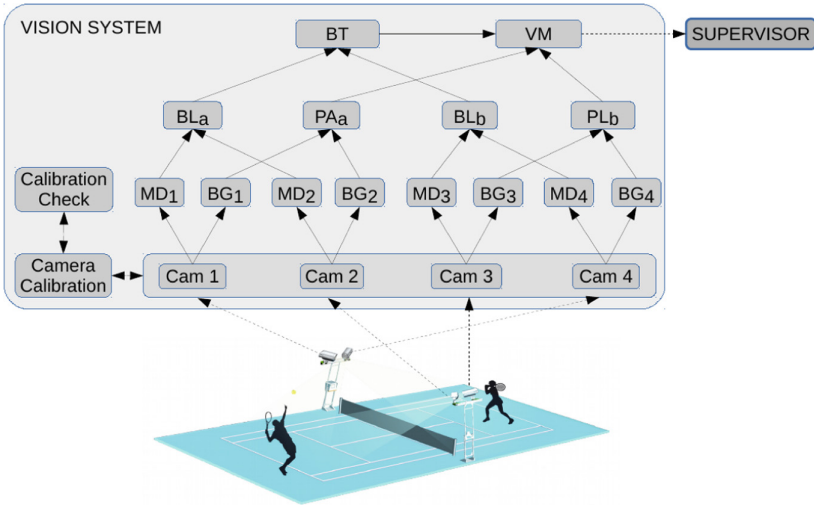


Fig. 2. Software architecture of the multi-thread Computer Vision system. The Vision Manager (VM) communicates with Supervisor and orchestrates the activity of the analysis modules: Motion Detection (MD), Background Analysis (BG), Ball Localization (BL), Player Localization and tracking (PL), and Ball Tracking (BT). Analysis modules use camera parameters produced and checked by two special modules. Each camera manager (Cam) acts as interface between a camera and the processing modules.

Configuration. The vision modules behaviour depends on the value of a set of parameters that have to be defined at installation time. The most relevant regard court and cameras: A COURT MODEL stores the position court elements (baselines, service lines, lateral lines, ...) with respect to a real-world coordinate system having origin in the centre of the court, X axis along the middle court line, Y axis along the net line and Z axis oriented upwards. The model includes also information about the court surface (clay, hard, grass, carpet). CAMERA CALIBRATION is a critical but essential step for 3D object localization from 2D images. Intrinsic camera parameters are estimated once and for all in laboratory by acquiring images of specific graphical patterns. Extrinsic parameters are estimated by an automatic module that works on images acquired at installation time after fixing cameras on their supports with a proper field of view. The auto-calibration module localizes the court lines intersections and put them in correspondence with the real-world positions stored in the COURT MODEL to calculate camera coordinates and orientation with respect to the reference system. An auxiliary module, called CALIBRATION CHECK, has been implemented for the periodical verification of camera poses. The alignment of expected court lines, according to calibration, and real ones is computed and if misalignment overcomes a given threshold a re-calibration step is automatically executed.

Image Analysis for Object Detection. Image acquisition and delivery to processing modules is managed by the Cam modules (see Fig. 2): each camera acquires

color images 1920×1200 at 50 fps, or 1920×1024 at 60 fps. Images are stored in circular buffers and provided upon request to the analysis modules. Two approaches are adopted to detect foreground objects from single images: background subtraction and motion detection (or frame difference). Four background updating modules (*BG*), running in parallel, get frames from camera managers and process them to create and maintain four images that store the appearance of the scene without foreground objects, e.g. players. To compute the so called foreground the current frame is compared to the background image. The background reference image is periodically updated (typically 3/5 times per second) in order to model scene changes, which otherwise would lead to the detection of false objects. Our method for background generation and updating is based on [13]. Motion Detector (*MD*) is another low-level processing module that gets frames from *Cam* and puts them in a short circular buffer. These are processed to identify pixels corresponding to moving objects in an image and to produce a motion map. It is analyzed to detect regions compatible with a moving ball using two methods: (i) after thresholding, extraction of connected components and morphological filtering, (ii) detection of peaked moving regions having a circular or elliptical shape. For each frame, the motion detector output is a set of candidate ball regions. The module is applied to the entire frame or to selected sub-regions according to the global state of the ball detection system stored in the Vision Manager.

3D Ball Localization and Tracking. This task is in charge to Ball Localization (*BL*) and Ball Tracking *BT* modules. BL is based on the analysis of the 2D candidate ball regions produced by two MD modules looking at the same half court. The type of analysis depends on the state of the tracking system, in particular we distinguish two states: (a) the system is searching for a moving ball (*search*), (b) the system is currently tracking a ball (*track*). In the first case both BL modules run in parallel looking for a moving ball in their respective half court. A BL collects candidate regions from a short sequence of consecutive frames and, through triangulation, provides a cloud of ball candidates in the 3D space. By means of graph analysis, 3D candidates collected through time are filtered to build a ‘tracklet’ consistent with the physical model of a flying ball [8]. When a BL detects a 3D ball the state of the system switches to *track* and the expected trajectory of the ball is estimated according to the motion model. If the system is in state *track*, the computation of the current ball location depends on the distance of candidate regions to their predicted position according to the expected trajectory. The system estimates the new ball position along with a confidence factor. In case of high confidence the new position is used to update the expected trajectory. A low confidence means a detection failure which is notified to BT for a proper management of the condition. BT monitors the output of the BL modules and, as soon as it is not consistent with the predicted trajectory, it performs a detailed analysis in order to identify the reason, that typically is included in the following list: ball went out the field of view, ball bounced, ball went in the net, ball touched the net chord, ball hit by a player. If the module recognizes one of these conditions it deals with it in a specific way, otherwise it labels the track as lost and the state is set again to *search*.

The management of the bounce event is particularly critical to support the *line calling* functionality. As described in [7], the ball cannot be assimilated to a material point with non-zero mass, but as a moving elastic sphere, that rolls and slides, and its contact with the ground is not a point but an area. For this, we use a bouncing model [2, 5, 6], and describe the impact region as an ellipsis. Accurate detection of the ball during the impact with a racket is a critical task, mainly because of fast movements and possible ball occlusions. For this reasons we infer the time and space coordinates of the impact point by analysing in-coming and out-going branches of the ball trajectory. BT returns to VM the ball coordinates at each time and the occurrences of mid-level events, like bounce, ball in net, impact with racket, net chord, along with the spatio-temporal coordinates they happened.

Player Detection and Tracking. Player Localization module *PL*, one for each half-court, aims to estimate player position and movement (Fig. 3). It computes two foreground maps from the output of BG modules related to the same half court. The module exploits a correspondence map that associates foreground pixels to lists of possible player locations in the real-world. This map is pre-computed in an off-line configuration phase by virtually placing a player model on a grid of locations on the court, projecting it on the image plane, and collecting the coordinates of changed pixels. The correspondence map permits to compute, at run-time, a discretized probability map of player location, in real-world coordinates, by accumulating the contribution of every foreground pixel. The probability maps are multiplied to combine information coming from the two points of view. To take into account spatio-temporal continuity, a prior probability map is considered that enforce the player location to be in a neighbourhood of the previous one. At the beginning, the prior map favours positions close to the baseline. The entropy of the probability map is used to determine if a player is in the court: high entropy values mean low probability of a person to be in the scene. PL modules run typically at 5–10 Hz and return to the Vision Manager information about presence and position of players in each half court.

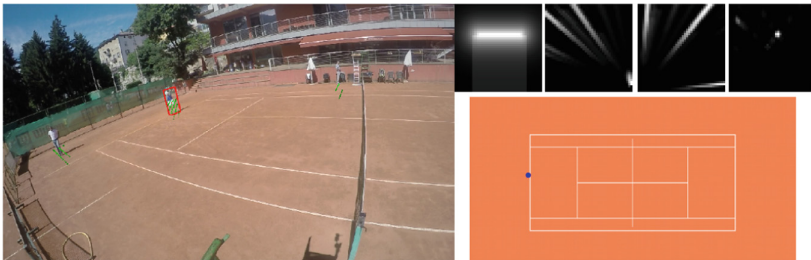


Fig. 3. Player localization: probability maps that explain foreground regions in two correspondent views (second and third maps) and an updated prior map of player location (first map) are used to estimates a posterior distribution (fourth map). The most probable location is reported in the 2D top view.

5 Analytics and Their Visualization

The Supervisor includes a module to process data produced by Vision System to compute video analytics, i.e. meaningful information about the observed activity useful for players to assess their skills and measure their progress. Knowledge about 3D trajectories of ball and players in real-time enables the extraction of useful data, like classification of the stroke type, building of bounce and serve maps, computation of shot speed, analysis of relative position of ball, player and racket during an impact. At an even higher level, it is also possible to obtain information about the sequence of individual strokes and to provide a description of player strategies. In the following we describe the most relevant implemented analytics.

Stroke Classification. In general, the stroke type depends on ball and player absolute and relative positions, e.g. serve, forehand, backhand, smash, but in order to provide a more detailed classification other data have to be considered: ball spin (top-spin, slice), ball direction (cross, long-line), opponent position (passing, lob), ball bounce before (volley) or after (drop-shot) the hit. At present, the system implements the classification of serve, forehand and backhand.

Line Calling. This analytic computes the possible intersection of the *legal region* with the ellipse provided by the Computer Vision along with the bounce event, to establish if the ball has fallen in or not. The *legal region* is differently defined by the tennis rule depending on the fact that the stroke originating the bounce is a serve, or not. In the first case it depends on the server player position (cross-court service box), while in all the other cases it is the opposite half court. The analytic considers therefore the playing situation: type of stroke and, if necessary, player position.

Other information are computed and made available to the players, like landing point map, player occupancy and movements maps, traveled distance, as well as statistics on speed/spin of the ball by type of stroke. Results of analytics are almost instantaneously accessible to players, through the on-court GUI. Data is visualized on the dashboard in a traditional tabular form as well as by means of geolocalized maps which enable the player to interpret spatio-temporal data at a glance. Figure 4 reports some examples. Video of the whole match or selected video clips are uploaded to a web server at the end of the game session and then accessible from any web device and shareable on social media.

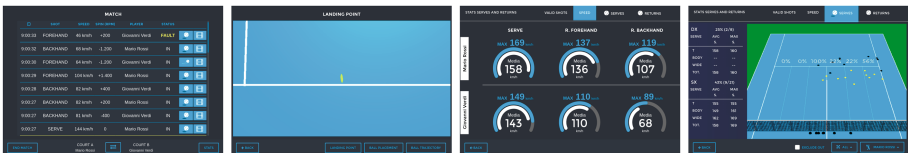


Fig. 4. Analytics examples. From left to right: real-time visualization during a match (shot classification, speed, spin); reconstruction of the landing point; statistics and map about a serve drill.

6 System Performance

EYES ON is currently installed in three tennis clubs and has almost completed the test phase (Fig. 5): Circolo Tennis (CT) Trento (Italy) - clay court covered with an air dome system during autumn and winter; Circolo Tennis Arco (Trento, Italy) - hard court inside a sport building; Centro di Preparazione Olimpica (CPO) of Tirrenia (Pisa, Italy) - belongs to the Italian Tennis Federation which has decided to evaluate EYES ON as a tool to support training of young top players. The various installations allowed players with different needs to test the system and provide their qualitative feedback. To assess the system quantitatively, we collected and examined data contained in a total of 1 h 23' 35" of game (match and drills), corresponding to 250.750 frames per camera, in sequences acquired from different court scenarios. In the considered videos 1069 shots are present, defining as shot the event in which a flying ball has an impact with a racket. The system has correctly detected 1061 shots (Precision 99.3%) and generated 3 false detection (Recall 99.7%). Table 1 reports performance about stroke classification by type (*Forehand*, *Backhand*, *Serve*) and by destination (*IN*, *OUT*, *Fault*). In the tables each row represents the instances of a true class, while the columns represent the classification provided by the system. In both cases data refers to the set of 1061 shots correctly detected. The system exhibits a reliability rate greater than 97% to declare the correct type of stroke and a reliability of 99.5% relatively to the shot destination.



Fig. 5. System installed at CT Trento court without and with covering. Camera on the timber joists at CT Arco, and cabinet at CPO Tirrenia.

Running Time. We evaluate system performance in terms of processing time through various working sessions, with different lighting conditions and game modes. Table 2-Left reports mean elaboration time for the most time consuming tasks, i.e. 3D ball localization and player localization. On average, half of the time is dedicated to the first task, while player localization, which is computed in parallel at a lower frame rate, requires only 75 ms every second. Therefore, the processing power can be dedicated to the other software components of the system for about 50% of the time, on average.

System Reactivity. We have estimated the *reaction time* of the system, that is the time needed by the system to signal an event (hit or bounce) after it happened.

Table 1. Left: Confusion table related to the classification of the three basic stroke types. Right: Confusion table related to the classification of shot as *IN*, *OUT*, *Fault*. GT columns report the ground-truth number of events.

	GT	<i>Fore</i>	<i>Back</i>	<i>Serve</i>	Reliability(%)	GT	<i>IN</i>	<i>OUT</i>	<i>Fault</i>	Rel(%)
<i>Forehand</i>	590	571	19	0	96.8	<i>IN</i>	855	852	3	99.7
<i>Backhand</i>	310	12	298	0	96.1	<i>OUT</i>	104	1	103	99.0
<i>Serve</i>	161	0	0	161	100.00	<i>Fault</i>	102	1	0	99.0
total	1061				97.1	total	1061			99.5

It is a sum of a small constant time, intrinsic in low-level image processing, plus the time to reliably detect the ball trajectory after the event and, finally, the time to compute the shot parameters. As reported in Table 2-Right complete information about the shot preceding the event is available after about 300 ms the event occurred. In the case spin calculation is not required, e.g. for in/out estimation, the average reaction time results to be 152 ms.

Table 2. Left: Average processing time for the main computer vision tasks: 3D ball localization and player localization. Right: Statistics on system reactivity (in milliseconds). We reported the average delay times (μ) in detecting a shot or a bounce, with and without computation of motion parameters (MP), and the standard deviation σ .

task	time (ms)	freq (Hz)	ms per sec	event	samples	with MP	w/o MP	σ
<i>ball</i>	10	50	500	<i>racket</i>	447	319	220	124
<i>player</i>	15	5	75	<i>bounce</i>	291	307	152	84

7 Conclusion

We have described a real-time vision-based system that offers to tennis players a new training and matching style. The system collects data and provide analysis usable by players, managers, coaches to improve game performance, highlighting weaknesses or strengths. The system is low-cost, flexible, easy to install, user friendly and reliable. Positive feedback has been collected from players who tested the system in three pilot installations. System reliability has been assessed in tests on real games. According to market needs, in the future we would like to extend the functionality of the system and improve its accuracy in order to submit the product to the ITF evaluation for obtaining the approval for automated line calling systems. Furthermore, research work continues to extend the system to analysis of other sports by taking advantage of the flexibility of the implemented modules.

References

1. Baodong, Y.: Hawkeye technology using Tennis match. *Comput. Model. New Technol.* **18**(12C), 400–402 (2014)

2. Brody, H.: Bounce of a tennis ball. *J. Sci. Med. Sport* **6**(1), 113–119 (2003)
3. Conaire, C.O., Kelly, P., Connaghan, D., O'Connor, N.: TennisSense: a platform for extracting semantic information from multi-camera tennis data. In: International Conference on Digital Signal Processing, Santorini, Greece (2009)
4. Connaghan, D., Moran, K., O'Connor, N.E.: An automatic visual analysis system for tennis. *Proc. Inst. Mech. Eng. Part P J. Sports Eng. Technol.* **227**(4), 273–288 (2013)
5. Cross, R.: The bounce of a ball. *Am. J. Phys.* **67**(3), 222–227 (1999)
6. Cross, R.: Measurements of the horizontal coefficient of restitution for a superball and a tennis ball. *Am. J. Phys.* **70**(5), 482–489 (2002)
7. Cross, R.: The footprint of a tennis ball. *Sports Eng.* **17**(4), 239–247 (2014)
8. Cross, R., Lindsey, C.: Measurements of drag and lift on tennis balls in flight. *Sports Eng.* **17**, 89–96 (2014)
9. In/Out: The Portable Ready-to-Use Line Call Device (2019). <https://inout.tennis>
10. ITF: Player Analysis Technology. PlaySight SmartCourt. Approval report PAT-13-003, Int. Tennis Federation, February 2014
11. ITF: Player Analysis Technology. FOXTENN. Approval report PAT-16-014, Int. Tennis Federation, May 2017
12. ITF: Player Analysis Technology. Hawk-Eye. Approval report PAT-18-001, Int. Tennis Federation, July 2018
13. Messelodi, S., Modena, C.M., Segata, N., Zanin, M.: A Kalman filter based background updating algorithm robust to sharp illumination changes. In: Roli, F., Vitulano, S. (eds.) *ICIAP 2005. LNCS*, vol. 3617, pp. 163–170. Springer, Heidelberg (2005). https://doi.org/10.1007/11553595_20
14. Owens, N., Harris, C., Stennett, C.: Hawk-eye tennis system. In: International Conference on Visual Information Engineering, Surrey, UK, pp. 182–185 (2003)
15. Pingali, G., Jean, Y., Opalach, A., Carlbom, I.: LucentVision: converting real world events into multimedia experiences. In: IEEE International Conference on Multimedia and Expo, Singapore (2000)
16. Pingali, G., Opalach, A., Jean, Y., Carlbom, I.: Instantly indexed multimedia databases of real world events. *IEEE Trans. Multimed.* **4**, 269–282 (2002)
17. Poliakov, A., Marruad, D., Reithler, L., Chatain, C.: Physics based 3D ball tracking for tennis videos. In: International Workshop of Content Based Multimedia Indexing, Grenoble, France (2010)
18. Renó, V., et al.: A technology platform for automatic high-level tennis game analysis. *Comput. Vis. Image Underst.* **159**, 164–175 (2017)
19. Teachabarikiti, K., Chalidabhongse, T., Thammano, A.: Players tracking and ball detection for an automatic tennis video annotation. In: International Conference on Control, Automation, Robotics and Vision, Singapore (2010)
20. Thomas, G., Gade, R., Moeslund, T., Carr, P., Hilton, A.: Computer vision for sports: current applications and research topics. *Comput. Vis. Image Underst.* **159**, 3–18 (2017)
21. Yan, F., et al.: Automatic annotation of tennis games: an integration of audio, vision, and learning. *Image Vis. Comput.* **32**, 896–903 (2014)
22. Yoo, W., Jones, Z., Atsbaha, H., Wingfield, D.: Painless tennis ball tracking system. In: IEEE Computer Software and Applications Conference, Tokyo, Japan (2018)