Computer Vision and Image Understanding 159 (2017) 3–18

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Computer vision for sports: Current applications and research topics

Graham Thomas[a], Rikke Gade[b], Thomas B. Moeslund[b,*], Peter Carr[c], Adrian Hilton[d]

[a] BBC R&D, 5th Floor, Dock House, MediaCity UK, Salford, M50 2LH, UK
[b] Aalborg University, Rendsburggade 14, 9000 Aalborg, Denmark
[c] Disney Research, 4720 Forbes Ave., Suite 110, Pittsburgh, PA, 15231, USA
[d] Centre for Vision, Speech & Signal Processing, University of Surrey, Guildford, GU27XH, UK

ARTICLE INFO

Article history:
Received 21 July 2016
Revised 27 February 2017
Accepted 22 April 2017
Available online 26 April 2017

Keywords:
Player tracking
Ball tracking
Sports analysis

ABSTRACT

The world of sports intrinsically involves fast and accurate motion that is not only challenging for competitors to master, but can be difficult for coaches and trainers to analyze, and for audiences to follow. The nature of most sports means that monitoring by the use of sensors or other devices fixed to players or equipment is generally not possible. This provides a rich set of opportunities for the application of computer vision techniques to help the competitors, coaches and audience. This paper discusses a selection of current commercial applications that use computer vision for sports analysis, and highlights some of the topics that are currently being addressed in the research community. A summary of on-line datasets to support research in this area is included.

1. Introduction

Computer vision already plays a key role in the world of sports. Some of the best-known current application areas are in sports analysis for broadcast, for example showing the position of players or the ball as 3D models to allow the locations or trajectories to be explored in detail by a TV presenter. Computer vision is also used behind-the-scenes, in areas such as training and coaching, and providing help for the referee during a game. Motion capture systems, relying on reflective cameras attached to athletes viewed by multiple cameras, are used in the training of professional athletes, and current research work is looking at how easier-to-deploy vision systems might be able to be used for similar tasks in the future. Other current research topics include analysis of how groups of players move, for applications like coaching in team sports, or automatically identifying key stages in a game for gathering statistics or automating the control of broadcast cameras. This paper presents an overview of how computer vision is currently being applied in sports, and discusses some of the current research that will lead to future commercial applications.

The remainder of this paper is organized as follows: Section 2 looks at how fundamental techniques such as tracking players and ball, and analyzing the motion of both individual players and teams, are being applied in today's commercial systems. Section 3 looks at how these techniques are being further developed, making reference to examples from recent publications, including others in this special issue. It also provides an overview of some publicly-available datasets to support ongoing research.

2. Current commercial applications of computer vision in sports

This section looks at how various fundamental techniques are being applied in commercially-available systems today. Applications that detect and track players and the ball are discussed, as well as those that track camera movement. Some current examples of techniques being used commercially to analyze the motion of players (both individually and within teams) are also briefly discussed, as are applications in enhancing sports broadcasts. Characteristics of some example systems are listed in tables at the end of the section.

2.1. Camera calibration and tracking

Camera calibration is essential for the ball and player tracking systems described in the following sections, and also for any systems that need to render graphics into the image that appear locked to the real world (or 'tied-to-pitch'). Those such as multi-camera ball tracking systems usually work with fixed cameras, and many calibration approaches can be used, including the use of calibration targets. Scene calibration may even use approaches such as rolling balls over the ground to account for non-planarity of the playing surface.

* Corresponding author at: Aalborg University, Laboratory of Computer Vision and Media Technology, Niels Jernes Vej 14 (3-109), DK-9220, Aalborg East, Denmark.
E-mail address: tbm@cvmt.aau.dk (T.B. Moeslund).

http://dx.doi.org/10.1016/j.cviu.2017.04.011
1077-3142/

Fig. 1. Camera with sensors on lens and pan/tilt mount for virtual graphics overlay.

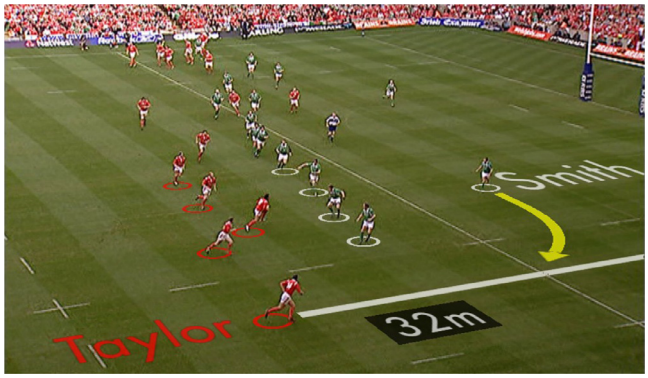


Fig. 2. Example of virtual graphics overlaid on a rugby pitch [picture courtesy of Ericsson].

Systems that work with broadcast cameras need to be able to account for changing pan, tilt and zoom. The first such systems relied on mechanical sensors on the camera mounting, and sensors on the lens to measure zoom and focus settings. Calibration data for the lens is needed to relate the "raw" values from these lens encoders to focal length. Ideally lens distortion and "nodal shift" (movement of the notional position of the pin-hole in a simple camera model, principally along the direction-of-view) should also be measured by the calibration process and accounted for when images are rendered. The position of the camera mounting in the reference frame of the sports pitch also needs to be measured, for example by using surveying tools such as a theodolite or range-finder. An example of a camera and lens equipped with sensors, for placing virtual graphics on athletics coverage, is shown in Fig. 1. However, this approach is costly and not always practical, for example if the cameras are installed and operated by another broadcaster and only the video feed itself is made available. The sensor data has to be carried through the programme production chain, including the cabling from the camera to the outside broadcast truck, recording, and transmission to the studio. Also, any system that relies on sensor data cannot be used on archive recordings for which no data are available.

Most current sports graphics systems that require camera calibration now achieve this using computer vision, using features at known positions in the scene. This avoids the need for specially-equipped lenses and camera mounts, and the problems with getting sensor data back from the camera. In sports such as soccer where there are prominent line markings on the pitch at well-defined positions, a line-based tracker is often used (Thomas, 2007). In other sports such as ice hockey or athletics, where lines are less useful, a more general feature point tracker can be used (Dawes et al., 2009). However, for live use in applications where there are very few reliable static features in view (such as swimming), sensor-based systems are still used. In particularly challenging cases where a stable camera mounting is not available and very high angular accuracy is needed (approaching 1/10,000 of a degree) to cope with zoom lenses having a very narrow field-of-view, it can be necessary to employ high-accuracy gyros. An example is the GyroTracker developed by Mo-Sys (Mo-Sys, 2016). For airbourne systems, both gyros and GPS have been used, for example in coverage of the America's Cup (Sportvision, 2016b).

The addition of some relatively simple image processing can allow graphics to appear as if drawn on the ground, and not on top of people or other foreground objects or players, if the background is of a relatively uniform colour. For example, for soccer, a colour-based segmentation algorithm (referred to as a "chromakeyer" by broadcast engineers) tuned to detect green can be used to inhibit the drawing of graphics in areas that are not grass-coloured, so that they appear behind the players. The fact that the chromakeyer will not generate a key for areas such as mud can actually be an advantage, as long as such areas are not large, as this adds to the realism that the graphics are "painted" on the grass. Fig. 2 shows an example of graphics applied to rugby; other examples of this kind of system applied to American Football include the "1st and Ten™" line (Sportvision, 2016a) and the "First Down Line" (Orad, 2016).

2.2. Detection and tracking

2.2.1. Player detection and tracking

Detecting the position of players at a given moment in time is the first step in player tracking, and is also required in sports graphics systems for visualization of key moments of a game. Techniques used in commercial broadcast analysis systems range from those relying on a human operator to click on the feet of players in a calibrated camera image (Bialik, 2014) to automated techniques that use segmentation to identify regions that likely to correspond to players (Tamir and Oz, 2008).

To help improve the performance of teams in sports such as soccer, analyzing the ways in which both individual players move, and the overall formation of the team, can provide very valuable insights for the team coach. Ideally, the coach would have access to a complete recording of the positions of all players many times per second throughout an entire training session or actual game.

Commercial multi-camera player tracking systems tend to rely on a mixture of automated and manual tracking and player labeling. STATS SportVu uses six cameras to track players in basketball; for football it uses a cluster of 3 HD cameras in single location, with an optional second cluster of 3 to provide height information (STATS, 2017). Sportvision uses a camera-based system for its FIELDf/x player tracking in baseball (McSurley and Rybarczyk, 2011). Academic groups are also developing multi-camera player tracking systems with the aim of creating commercial products, such as such as the Computer Vision Laboratory at EPFL and its spin-out PlayfulVision (EPFL, 2016).

Fully-automated tracking and labelling of players remains an open challenge. Optical tracking systems must cope with players occluding each other and having similar appearance. Some recent research in this area is discussed in Section 3.2.

2.2.2. Ball tracking

The ability to track a ball in low-latency real-time is important for both analysis in broadcast TV and helping the referee or umpire. One of the first commercially-available multi-camera systems based on computer vision was developed by Hawk-Eye (Innovations, 2017c) for tracking cricket balls in 3D, and first deployed in 2001. It was subsequently applied to tennis; al-

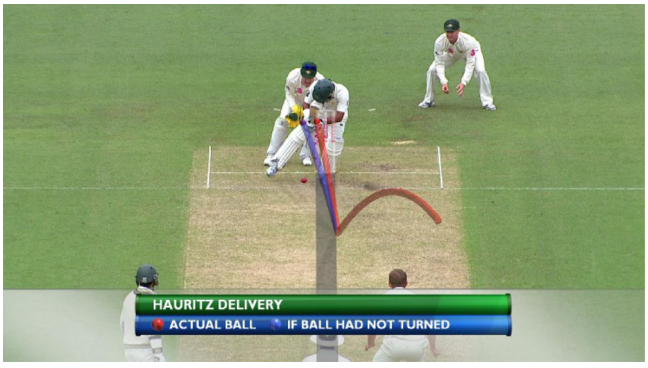


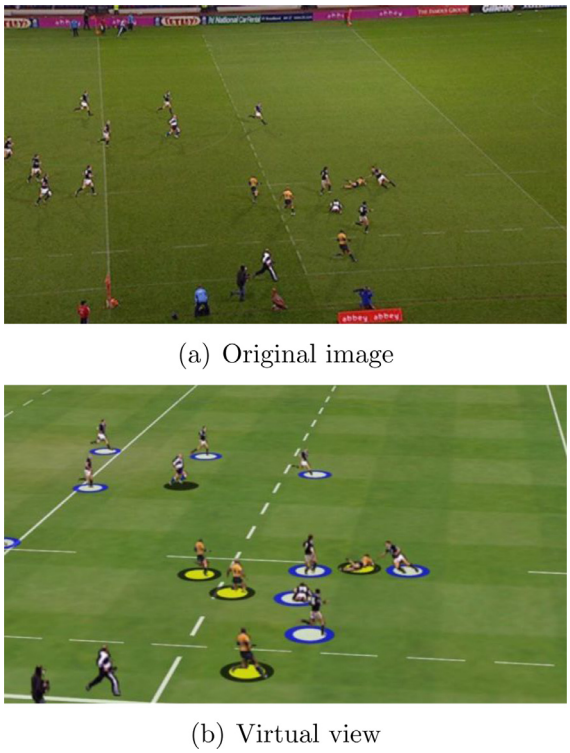
Fig. 3. Analysis and prediction of cricket ball motion [picture courtesy of Hawk-Eye Innovations].

though broadcast cameras were initially used to provide the images (Owens et al., 2003), the system is generally now deployed with up to ten cameras set up around the court to capture live images at up to 340 fps. (Innovations, 2017a). Being static, they are easier to calibrate, and short shutter times and higher frame rates can be used. There is a lot of prior knowledge about tennis that the system can use, including the size and appearance of the ball, its motion (once it is hit, its motion can be predicted using the laws of physics), and the area over which it needs to be tracked. The cameras and the geometry of the court can be accurately calibrated in advance. The system first identifies possible balls in each camera image, by identifying elliptical regions in the expected size range. Candidates for balls are linked with tracks across multiple frames, and plausible tracks are then matched between multiple cameras to generate a trajectory in 3D (Owens et al., 2003). The system is sufficiently accurate that it can be used by the referee to determine whether a ball lands in or out, being first used for this in 2005. It is claimed to have a mean error of 2.6 mm when compared to a high speed camera located on the playing surface (Innovations, 2017a). When applied to cricket, it typically uses six cameras running at 340 fps and achieves an accuracy of 5–10 mm (Innovations, 2017a). It can be used to predict the path that the ball would have taken if the batsman had not hit it, or if it had not turned when bouncing, as shown in Fig. 3. A similar system has been used for officiating in badminton by the Badminton World Federation since 2014 (Badminton, 2017).

Sportvision's PITCHf/x system (Fast, 2010) tracks a baseball using a pair of cameras, and was first deployed in 2006. The real-time data was used for ESPN's K-Zone visualization technology. In 2015 (Cole, 2014), Major League Baseball investigated radar-based ball tracking and optical-based player tracking using ChyronHego's TRACAB (ChyronHego, 2017b).

In 2006, Protracer introduced a vision-based golf ball tracking system using a specialized image sensor (Anderson, 2013). ESPN used the technology to generate visualizations over live broadcasts in 2010 (Hall, 2014).

More recently, various systems to determine whether a soccer has crossed the goal line have been developed. At the time of writing (Feb 2017), FIFA has certified goal-line technology (GLT) installations at 108 soccer grounds (FIFA, 2016a). Of these, 87 are based on a 7-camera computer vision system developed by Hawk-Eye, and 21 use a similar vision system developed by GoalControl (FIFA, 2016a). The GoalRef system (IIS, 2016) (an RF system with a transponder in the ball) developed by Fraunhofer IIS is also certified by FIFA. The system from GoalControl was used for the 2014 World Cup in Brazil (FIFA, 2016c), and the system from Hawkeye was used for the Euro 2016 championships in France (UEFA, 2016).



(a) Original image

(b) Virtual view

Fig. 4. Generation of a virtual view from a single camera image [picture courtesy of Ericsson].

2.3. Broadcast enhancements

2.3.1. Player modeling

Once players have been located, various approaches are used in current sports graphics systems to provide a visualization of the key moment from a range of viewpoints.

An early example of the use of multiple cameras to provide a view from a range of points around an area of interest was the EyeVision system developed by CBS with help from Carnegie Mellon University for the Super Bowl in 2001 (Mellon University, 2017). 30 cameras with motorised zoom and focus on robotic pan/tilt heads followed a viewpoint on the pitch defined by an operator. A video switching system switched quickly between the views to give the impression of movement around the point of interest.

Computer vision techniques have subsequently been used to simulate a smoothly-moving viewpoint from one or more fixed cameras. A crude 3D model of the scene can be created from a single camera using a billboard approach (Grau et al., 2001) by segmenting the players from the background and placing them into a 3D model of a stadium, as textures on flat planes positioned at the estimated locations. An example of this approach is shown in Fig. 4 produced by Ericsson (2016); this was first used in 2004 (R&D, 2017). This allows the generation of virtual views of the game from locations other than those at which real cameras are placed, for example to present a view of the scene that a linesman may have had when making an offside decision, or to provide a seamless "fly" between an image from a real camera and a fully-virtual top-down view more suitable for presenting an analysis of tactics.

This simple player modeling approach works well in many situations, but the use of a single camera for creating the models restricts the range of virtual camera movement, with the planar nature of the players becoming apparent when the viewing direction changes by more than about 15 degrees from that of the orig-

inal camera. Furthermore, overlapping players cannot easily be resolved. Approaches using two or more cameras, to create either a "2.5D" model that allows smooth blending between multiple billboarded images, or to infer the player pose in 3D by matching player silhouettes to a library of pre-generated poses, are described in Popa et al. (2014), and techniques like this are available in commercial products such as Viz Libero (Vizrt, 2016).

The use of a large number of cameras allows a wider range of movement and more accurate models to be created. Such a system (Technologies, 2017) was used at the 2012 Olympics for gymnastics, and then under the name EyeVision 360 at the 2016 Superbowl, incorporating 36 cameras (Takahashi, 2017). It has also been used for other sports including tennis and basketball (Petrovic, 2017).

An alternative solution to these problems is to use pre-generated 3D player models, manually selected and positioned to match the view from the camera. It can take a skilled operator many minutes to model such a scene, which is acceptable for a post-match analysis programme but too slow for use in an instant replay. The player models lack realism, but can be at a level of detail higher than can readily be achieved by a camera-based system.

2.3.2. Analysing motion of players

Analysing or visualising the motion of players at key moments in sport can give useful insights for both trainers and broadcasters. Training at elite level may involve the use of dedicated motion capture systems with multiple calibrated cameras and markers placed on the player. However, such marker-based systems are impractical to use in lower-end training sessions and during actual competitions. In these situations, computer vision using a single camera or a small number of cameras can still provide useful insights. Most commercial systems are currently limited to providing visualisation of the player's movement rather than detailed analysis, as accurate markerless pose estimation is very challenging.

The motion of the segmented foreground person or object can be illustrated by overlaying a sequence of "snapshots" on the background, to create a motion "trail', allowing the recent motion of the foreground object or person to be seen. This produces a similar effect to that which can be achieved by illuminating the scene with a stroboscope and capturing an image using a camera with a long exposure. An early example of this was an analysis tool for snooker (Storey, 1984) which exploited the relatively benign nature of this sport (cameras are often static, and balls are easily segmented from the plain green background).

A more recent example of this kind of analysis tool (Prandoni, 2003) is often used to show movement in sports such as diving or ice skating. It is necessary to compensate for any pan/tilt/zoom of the camera, so that the segmented snapshots of the people are shown in the correct place; this can be achieved using image analysis techniques like those discussed above. This application is relatively benign to segmentation errors which might result in the inclusion of parts of the background around the edges of a segmented person, as the person is overlaid on the background region from which they were originally extracted. By stitching together the background areas from a sequence of images and viewing this as one large image, it is possible to illustrate the entire movement of a person over a given period of time, as shown in Fig. 5. Where the person tends to occupy the same location over a period of time, such as an ice skater spinning around her axis, a set of successive images can be separated out and displayed on an arbitrary background, like a series of frames in a film.

An extension of this class of technique can be used to overlay the performance of several sports people during successive heats of an event, such as a downhill ski race (Reusens et al., 2007). This requires the calibration of the moving camera in a repeatable way, so that the background can be aligned from one heat to the next,

Fig. 5. An illustration of motion using a series of snapshots of the action [picture courtesy of Dartfish].

and may also require the timing to be synchronised so that competitors are shown at the same time from the start of the heat.

Recently, machine learning and data mining techniques have been used to better understand raw player tracking data in team sports. To some degree, the methods of analysis that can be used are somewhat dependent on how the data was acquired. For instance, the noise characteristics of optical tracking are quite different from those of radio frequency tracking. As a result, we only highlight a few recent examples of player tracking analysis, since this topic is mostly outside of the scope of computer vision.

Labelling semantic events, such as "pick and roll" in basketball, is a popular topic when analysing player tracking data (McQueen et al., 2014). These labels enable broadcast enhancements such as advanced statistics (*e.g.* estimated shot quality Chang et al., 2014) or faster querying of video archives via keywords. In some cases, the analysis is used as a coaching tool, such as "ghost players" which illustrate a model's predicted optimal positions for players (Lowe, 2013).

2.4. The impact of computer vision technologies on sport

The use of computer vision and associated graphics technologies to provide insight and explanation into sports events has been a part of the evolution of sports TV, building on older techniques such as the slow-motion replay. These developments have been largely embraced by viewers and broadcasters. However, from the time that slow-motion replay first became available; it has allowed refereeing decisions to be examined in detail, inevitably leading to debate. Examples of comments made in the context of football include (Stack Exchange, 2017):

- Uncertainty is a part of the game. The referee and linesmen are the arbiters of the action. Instant replay would limit their authority.
- Football is global. Many places do not have access to the technology necessary to implement instant replay.
- Football is a fast-paced game with few opportunities for stoppage. Instant replay would unnecessarily slow the pace.

A Video Referee has been used in the National Basketball Association since 2002 (Broussard, 2017), following cases where the referee could clearly have been seen to make a mistake by looking at replays. Similar pressure led to FIFA approving the use of goal-line technologies in 2012 (Stack Exchange, 2017), and the debut of the Hawk-Eye tennis system for line calls at the 2006 US Open has changed the dynamics of professional tennis (Fetters, 2012).

Player tracking technology has revolutionised training and scouting for players in sports such as football and basketball. Some insights into how the data is used in basketball can be found in Aldridge (2013), including comments that players appreciated being shown data to back up what the coach was telling them, but also that there are fears that statistics could be used against players when negotiating contracts.

2.5. Player tracking using alternatives to computer vision

For some sports, tag-based RF tracking technology (such as GPS) has been used as an alternative to vision-based systems, for ex-

Table 1
Examples of some commercial systems: camera tracking, overlays and virtual views.

Manufacturer & system	Primary application area	Camera configuration and approach	Speed of operation and degree of autonomy	Precision and performance characteristics	How widely used
Ericsson: Piero (Ericsson, 2016)	Overlay graphics for TV and virtual views using either fully-virtual or image-based player models	Single pan/tilt/zoom broadcast camera under control of cameraman, no additional sensors, uses lines (Thomas, 2007) and/or point features (Dawes et al., 2009)	Camera tracking: automatic (latency of a few frames); Initialisation: semi-automatic or automatic (Thomas, 2007); Player modelling: billboards (Grau et al., 2001) or manually-posed 3D player models	Tracking stability generally better than 1 pixel	As of 2011, Piero was in use in over 40 countries. At the 2010 World Cup in South Africa over 85% of all analysis effects used in the world feed came from Piero (R&D, 2017). Other similar systems include (Orad, 2016), (Sportvision, 2016a) and (Vizrt, 2016).
Mo-Sys: Gyro tracker (Mo-Sys, 2016)	Tracking system to drive overlay graphics for TV for Red Bull air race	Inertial sensors mounted on a single camera (no computer vision used)	Assumed to be very low	Nearly 10,000th of a degree (Mo-Sys, 2016)	Specifically developed for Red Bull air race
CBS/CMU: Eye vision (Mellon University, 2017)	Moving viewpoint in a ring around an area of interest	30 cameras on robotic pan/tilt mounts at intervals of 7 degrees. No view interpolation.	Focus of interest set by an operator. Replay available virtually instantly.	No information readily available	Used for Superbowl XXXV broadcast, 2001. Apparently not used on subsequent Superbowls.
Replay Technologies (Intel) FreeD (Technologies, 2017)	Free-viewpoint video for sports	20–40 5K high-definition cameras (fixed mountings). Foreground and background modelled separately (Haimovitch, 2015)	Produces results quickly enough to use for replays in-venue or on TV (no info readily available on actual processing time)	No information readily available	First used for gymnastics at 2012 Olympics. First Superbowl use in 2016 (36 cameras, branded as EyeVision 360 (Takahashi, 2017)). Used for basketball (3 stadiums fitted with 28 cameras for the 2016 NBA All-Star game). Also used for tennis and baseball (Petrovic, 2017).

ample tags using GPS positioning and including heartrate monitors have been used for rugby, specifically for analyzing physiological demands on players, fatigue during game, and rehabilitation and injury prevention (Gilligan, 2014). Tag-based systems can suffer from problems like short battery life (reportedly under 4 h for some systems) and potential interference problems in bad weather. Tag-based systems are not restricted to outdoor scenarios where GPS signals can be received; various systems are available that use in-stadium transponders to transmit or receive reference signals. The need to get players to wear tags can meet with resistance, for example wearables are currently barred from being used during NBA games, and the data gathered from them cannot be used in NBA player contract negotiations (Leung, 2017).

2.6. Summary of some commercial systems

The tables below summarise some of the commercial systems referred to in this section. Table 1 focuses on systems primarily providing enhancements for viewers, using technology including camera tracking and view synthesis. Table 2 gives some details of player tracking systems, used for both coaching and broadcast statistics. Table 3 summarises some systems for high-precision ball tracking for both officiating and broadcast enhancements.

3. Open issues and current research areas

To fully automate the video analysis of sports events many issues are still open for research. The numerous different aspects of a sports event start when deciding how to capture the event; modelling the pose and viewpoint of a camera and calibrating large camera networks is still very challenging. There is then the key challenge of detecting and analyzing humans. Sports videos range from close-up views of single athletes to wide views including spectators at large stadiums, presenting a wide range of different scenarios. In this section we will discuss the largest research areas at the moment, which include detection and tracking of players and balls, as well as extracting semantics from sports videos. Lastly, we will give an overview of publicly available datasets useful for improving and comparing the research in these topics.

For sport as a human centred activity, the first step in most automatic analysis is to locate and segment each person of interest, and possibly follow this person over the duration of the video. Several challenges influence these tasks. The body posture of a person can vary greatly during sports exercises, decreasing the performance of any standard human/pedestrian detector. Another significant challenge is occlusion. People can be partly or fully occluded by, e.g., equipment, obstacles, or other players. In any contact sport or team sport occlusion between people is a frequent problem and includes cases of collisions and interactions between several players simultaneously.

3.1. Detection

3.1.1. Detection of players

The choice of detection methods depend on the type of footage, such as moving vs. static cameras, single vs. multiple cameras, static vs. changing background, and degree of occlusion.

Using static cameras it is possible to build a background model and detect players with simple methods like image differencing and background subtraction (Reno et al., 2015). On courts with uniformly coloured surfaces similar approaches can be applied with moving cameras using a colour-based elimination of the ground (Rao and Pati, 2015). These methods are often fast and well-suited for real-time performance. However, noise and missed detections

Table 2
Examples of some commercial systems: player tracking.

Manufacturer & system	Primary application area	Camera configuration and approach	Speed of operation and degree of autonomy	Precision and performance characteristics	How widely used
STATS SportVU (STATS, 2017)	Player and ball tracking for sports analytics for coaching/ training and broadcast, best known for basketball	For basketball (NBA): 6 cameras used; for football: a cluster of 3 HD cameras in single location, with an optional second cluster of 3 to provide height information. Some technical details in Tamir and Oz (2008).	Real time operation at 25Hz. Data can be augmented with play-by-play events from the official event feed (STATS, 2017).	For basketball: The data is tracked within 2–3 secs of capture; spotters (operators) can make changes to the assigned player labels. Sometimes, in scrums where several players are clustered, even though the player's heads remain visible, the system can temporarily lose track of who's who. Aldridge (2013).	Used by all 30 NBA teams (Wolverton, 2017)
ChyronHego TRACAB (ChyronHego, 2017b)	Player and ball tracking for coaching/training and broadcast in sports including football, tennis, basketball, baseball, cricket and American football	Two clusters of 3 HD cameras	Two operators assign player labels. Delivers "live tracking of all moving objects with a maximum delay of just three frames" (ChyronHego, 2017b).	Delivers "positioning accuracy across the whole field of play of less than the width of a hand" (ChyronHego, 2017b).	Installed in 125 arenas and is used in more than 2000 televised matches and games per year around the world (ChyronHego, 2017a). Employed across entire leagues such as the English Premier League. Details of approach to ball tracking (as of 2003) in Owens et al. (2003).German Bundesliga, Spanish La Liga and Major League Baseball, as well as some of the largest international sports tournaments including UEFA Champions League and FIFA World Cup.

Table 3
Examples of some commercial systems: ball tracking.

Manufacturer & system	Primary application area	Camera configuration and approach	Speed of operation and degree of autonomy	Precision and performance characteristics	How widely used
Hawkeye Cricket system (Innovations, 2017c)	Officiating in cricket, and broadcast enhancement	Six 340 frames per second cameras (Innovations, 2017e)	Real time, autonomous	5–10 mm accuracy. Ball size in image is circa 10 pixels, and the centre is found to circa 1/3 pixel (Innovations, 2017e).	Used by host broadcasters at major Test, ODI and Twenty20 matches around the world since 2001 and in 2008 was approved for use by the ICC and added as part of the Decision Review System
Hawkeye Tennis system (Innovations, 2017d)	Officiating in tennis, and broadcast enhancement	Up to ten cameras set up around the court to capture live images at up to 340 fps. (Innovations, 2017a). Details of approach to ball tracking (as of 2003) in Owens et al. (2003).	Real time, autonomous	Mean error of 2.6 mm when compared to a high speed camera located on the playing surface (Innovations, 2017a).	First used in 2002 as part of the BBCs Davis Cup coverage. First used for officiating in 2005. Now used in over 80 tournaments around the world and at all Grand Slam events. Innovations (2017a).
Hawkeye Goal-line system (Innovations, 2017b)	Officiating in football	7 cameras per goal, most commonly on the roof of the stadium.	Real time, autonomous	FIFA requirements include a positional accuracy of +/-1.5 cm and that an indication of a goal is provided in 1 s or less, as well as immunity to snow and toilet rolls (FIFA, 2014).	Used by the English Premier League, German Bundesliga, Dutch Eredivisie, Italian Serie A and FIFA

should be expected due to other moving objects, similar colours in foreground and background, and effects such as changing lighting conditions and shadows. Many methods for fine tuning and post-processing of foreground images have been proposed and it remains a challenge open for research (Gade et al., 2012; Reno et al., 2015).

Applying trained classifiers has become the most popular strategy for pedestrian detection. Naturally, as both are focused on detection of humans, this approach has been transferred to detection of sports players as well. However, sports players generally have much larger variations in pose, and the view-point and distance might change significantly between pedestrian and sports videos. Using the AdaBoost algorithm for training a classifier with HOG features, Faulkner and Dick (2015) shows that it is important to train the system with samples from the same video types, in this case Australian Rules Football, as a detector trained on standard pedestrian images performs very poorly on sports data. A similar approach with AdaBoost and Haar features is presented for detection of basketball players (Ivankovic et al., 2012), but the conclusion here is that although reasonable detection rates are obtained, the false positive rate is too high for the method to be applicable.

In team sports, additional steps have to be introduced to ensure the consistency in identity of the detected players. One approach is to solve the assignment task with classification of each object based on unique appearance features, like the jersey number (Gerke et al., 2015; Ye et al., 2005). Since these features are often only visible in a subset of frames, the classification can be combined with temporal tracking of the player to keep track of the identity (Lu et al., 2011). This topic will be further discussed in Section 3.2.

3.1.2. Detection of the ball

Many sports types have the game centered around a ball, which makes automatic detection of a ball interesting for classification of goals and other events, or statistics like ball possession. As described in Section 2.2.2, computer vision based goal detection systems are now developed to a point where these have been commercialised and trusted even for World Cup games. However, these systems utilise multiple view high-speed cameras covering each goal area, which means that extending a similar ball tracking system to the entire field would require an unrealistically large number of calibrated cameras. Detection and tracking the motion of the ball is particularly challenging due to fast motion and the small size of a ball compared to both humans and field sizes in most sports types.

Most ball detection algorithms get reasonable results by detecting moving objects and sorting these ball candidates on area, colour and shape (Chakraborty and Meher, 2012; Halbinger and Metzler, 2015; Wu et al., 2006). However, often a ball is partly or fully occluded by the players, in which case a pure detection algorithm will not be useful. Instead, combining detection and tracking might allow for synthesising the ball position based on the trajectory. This will be further discussed in Section 3.2.

3.2. Tracking

3.2.1. Tracking players

Tracking is one of the largest research areas of computer vision, with hundreds of papers published each year. Compared to other applications of human tracking, tracking of sports players is particularly challenging due to fast and erratic motion, similar appearance of players in team sports, and often close interactions between players is part of a game. Many multi-purpose tracking algorithms assume linear motion resulting in poor performance in sports compared to regular surveillance videos. Multi-target trackers commonly solve the data association problem with appearance models. However, this may fail in team sports due to the ambiguity of appearance between players on each team. In this section we will give an overview of some of the methods focusing specifically on solving the challenges of tracking in sports videos.

Recent research has investigated the use of context information, assuming that players react to the current game situation as a group. In Liu and Carr (2014) a method was presented using this approach by formulating a number of game context features, which significantly improved the tracking results on field hockey and basketball datasets. Context information has also been applied for tracking players in American Football (Zhang et al., 2012), and for soccer and volleyball (Xiao et al., 2014).

Sports games can also be modelled as a constrained closed-world scenario, in terms of both physical space and players" movements (Kristan et al., 2009). The severe challenges with occlusions observed in team sports have been addressed by a number of papers. Figueroa et al. (2006) use a graph representation, which includes the number of people assumed to be included in each detected blob. A different solution to occlusion handling is adding more cameras viewing the scene from different angles (Kasuya et al., 2008; Xu et al., 2004).

Many sports videos originate from broadcasts, which are interrupted by commercials, replays, and changing camera view. Consistent tracking of individual players across these interruptions is difficult due to the similar appearance of players. Soomro et al. (2015) propose a method to associate the tracks between video clips by modelling the team formation as a graph with each node representing a player position.

Manafifard et al. provides a detailed survey on player tracking in soccer videos (Manafifard et al., 2016).

3.2.2. Tracking the ball

Constructing trajectories of a ball is often of interest, e.g., for analysing shot types, play dynamics, and event detections in many sports types, or for virtual replays. Many proposed methods are physics based, starting with a set of ball detections, for which a physical model is fitted and the trajectory can be constructed and extended (Chakraborty and Meher, 2013; Chen et al., 2009; Leo et al., 2008).

In sports like tennis, where the ball is fully visible in most frames, it is possible to construct tracklets, which can be connected in a later association step (Yan et al., 2014).

Tracking the ball is harder in team sports, where several players can occlude the ball, and it is possible that players are in possession of the ball, either in their hands or between their feet. By combining knowledge on the positions of players and ball, the possession of the ball can be modelled and used as part of the ball trajectory (Wang et al., 2014; 2016; Wei et al., 2015). Maksai et al. (2016) continues this approach by defining a number of states the ball can take, such as *flying, rolling*, or *in_possession*. For each state specific constraints apply, which for the *flying* and *rolling* states are based on physics.

As opposed to players, the ball cannot be assumed to move on the ground plane, but rather in an unconstrained 3-dimensional space. To solve ambiguities often a setup of multiple calibrated cameras is applied (Ren et al., 2009) or even a stereo setup or RGB-D sensor for estimating depth in smaller areas, like a table tennis environment (Tamaki and Saito, 2014).

3.3. Semantics

3.3.1. Semantics from video

With the rapidly-growing amount of video data being captured and published, methods are sought for automatically labelling the data. This can be done by extracting the semantics from visual features.

Labelling can be divided into two levels of detail, with the first level being the sports genre, or in other words the general activity observed during a video clip. The second level is the individual actions performed, where one activity can consist of several different actions. Action recognition is often coupled with localization of the action in both space and time. Combining the two levels of categorisation, Wilson et al. (2014) detect events during video clips and use those for classification of the activity.

Activity recognition is often performed for an entire video clip. In these cases visual information about the environment, such as court colour and lines, can be extracted and used for classification (Krishna Mohan and Yegnanarayana, 2010; Mutchima and Sanguansat, 2012; Yuan and Wan, 2004). For general-purpose facilities, like public indoor arenas, where the environment does not provide information on the activity type, it has been investigated how to use only the information of players' positions over time. This data can originate from both visual tracking, but also other tracking technologies. From position data occupancy heatmaps (Gade and Moeslund, 2014) or trajectories (Lee and Hoff, 2007) can be constructed and used for classification.

Action recognition in videos can be used for automatic generation of statistics, like shot type, and for indexing of videos for easier browsing or generating summaries. The purpose of generating highlights or summarization of games is to shorten recordings of full-length games while still preserving the most interesting content, for example for TV news or personalized multimedia content on request (Chen et al., 2011). Summarizations are often constructed by event detection and classification (e.g., goals scored) (de Sousa et al., 2011; Zawbaa et al., 2011) or cinematographic features (e.g., camera motion and shot type) (Ekin et al., 2003; Nguyen and Yoshitaka, 2014). The most interesting events to detect during a game are often goals or score, but events such as penalties, near misses, and shots can also be of interest. Kapela et al. (2014) propose a method for detecting all of these interesting events in field sports, by including automatic interpretation of the scoreboard, and scene analysis based on the shot sequence.

Highlights can also be detected by the excitement level of the crowd, extracted from audio or audio-visual cues (Boril et al., 2010; Hanjalic, 2005; Kolekar and Sengupta, 2015).

Analysing the actions of individual sports types can also be used for performance analysis. For combat sports, Behendi et al. (2016) demonstrated the possibility of tracking boxers and classifying punches using RGB and depth data.

3.3.2. Semantics from data mining

Digging a little deeper into the activities performed at the sports field, analysing the motion of both individuals and teams can reveal information about behaviour and performance. These methods are related to data mining, and the data can originate from different tracking technologies, like radio frequency systems, GPS, or computer vision. In this section, however, we will only highlight methods based on video data.

Recognizing team behaviour, such as specific formations and play types, can be used for informing a tracking system, or for extraction of specific events. Bialkowski et al. (2014) use an occupancy map representation of team behaviour in field hockey with team classifications for each detection. Similarly, Atmosukarto et al. (2014) use players' positions to detect team formation, line of scrimmage and formation frame in American football videos. Fu et al. (2011) show that the screen strategy in basketball can be detected based on players' trajectories. Likewise, Perse et al. (2009) present a trajectory based method for analysis of team activities in basketball. However, obtaining reliable trajectories is still a challenge, as discussed in Section 3.2, and the work presented in Perse et al. (2009) is based on only semi-automatic tracking results.

3.4. Public datasets

In this section we present an overview of publicly available sports video datasets for which annotations are available. Shared datasets make it possible to compare the performance of different algorithms directly on the same data and improve the transparency of the research in the field. Furthermore, the extremely time-consuming process of capturing and annotating large quantities of diverse video can be reduced by sharing the data among researchers.

Table 4 provides an overview of some datasets and their characteristics. The datasets can be considered in two main categories: Still images or video, typically from moving cameras, of single athletes, with the purpose of activity recognition, and team sports videos, often captured with several static cameras, for tracking and event detections. One dataset focuses on the pose and actions of spectators of a sports event rather than players.

The datasets available for activity and action recognition are large with a great variety of sports actions. Most of these datasets are compiled from video sources like YouTube or mobile apps, hence unconstrained in view, scale, camera motion, etc. Varying from 150 to more than 1 million sequences per dataset, the amount of data is huge, but this reflects the variation and the number of different actions at the most fine-grained level, e.g., the Sports-1M Dataset with 487 activity classes.

The datasets of team sports activities still need additions to represent the variety between teams, environment, etc. The reason for the more limited quantity of video in this category may be partly because of the complex camera setup needed to cover the entire field, as well as the duration needed to represent different game events and transitions. Access to data is also restricted by rights issues for broadcast coverage of many sports events. Datasets are now available representing basketball, handball, soccer and squash with a few sequences per activity, sufficient to evaluate algorithms for tasks such as multi-target tracking. However, for approaches including game context or analysing sports-specific events, larger quantities of data from the same sports activity are needed.

The remainder of this section will describe each dataset and present sample images.

- **APIDIS Basketball Dataset**
  The APIDIS Basketball Dataset (Vleeschouwer et al., 2008) provides 16 min of basketball video captured with 7 cameras around and above a basketball court. Example frames from four views are shown in Fig. 6. The dataset is manually annotated with basketball events for the entire game, and positions of players, referees, baskets, and ball for one minute. Measurements and calibration images are provided for calibrating each camera to the common world coordinate system.
  The dataset is available for download from http://sites.uclouvain.be/ispgroup/index.php/Softwares/APIDIS.

- **CVBASE '06**
  The CVBASE '06 Dataset (Pers et al., 2006) provides three video subsets suitable for tracking and team/individual activity recognition. The first subset is a 10 min recording of European (team) handball. Video is available from three synchronized cameras: Two overhead cameras and one side-view (manually operated) camera. Trajectories are available for 7 players for all 10 min, in court coordinates and in coordinate systems of both overhead cameras. Furthermore, team activities (offense, defense, etc.) and individual activities (pass, shot, etc.) are manually annotated by a sport expert.
  The second subset is recordings of 2 squash matches (9 and 10 min duration). Trajectories are available for both players in court and camera coordinates. Furthermore, manual annotations include phases (rallies and passive phases), stroke type

Table 4
Characteristics of the datasets referred to in this section.

Dataset	Sports	Moving cam.	Annotations	Number of seq.	Avg. length	Quality
APIDIS Basketball Dataset	Basketball	No	Player trajectories Ball trajectory Event detections	1	16 min	1600 × 1200 pix, 22fps 7 Cameras
CVBASE '06	Handball	1 moving 2 static (overhead)	Trajectories Team activities Ind. activities	1	10 min	384 × 288 pix, 25fps
	Squash	No (overhead)	Trajectories Game actions	2	9m 46s	384 × 288 pix, 25fps
	Basketball	No (overhead)	NA	1	5 min	368 × 288 pix, 25fps
ISSIA Dataset	Soccer	No	Player trajectories Ball trajectories	1	2 min	1920 × 1088 pix, 25fps 6 synchronized cameras
Leeds Sports Pose Dataset	8 types	Still images	Joint locations	2000	Still images	Varying
Olympic Sports Dataset	16 types	Yes	Activity labels	800	NA	YouTube
S-HOCK dataset (spectators)	Hockey	No	Spectator position Spectator head pose Spectator posture Spectator action	15	31 s	30 fps
Soccer Video and Player Position Dataset	Soccer	No	Trajectories Ball position (subset)	5	45 min	High res. panorama and three single views
Sports Videos in the Wild	30 types	Yes	Activity label Action label Action localization	4200	15.1 s	480 × 270 pix or 480 × 360 pix, 30 fps
Sports-1M Dataset	487 types	Yes	Activity labels	1,133,158	5 m 36 s	YouTube
UCF Sports Action Dataset	10 types	Yes	Bounding boxes Activity labels Human gaze (viewer)	150	6.3 s	720 × 480 pix, 10fps
UIUC Sports Event dataset	8	Still images	Activity labels	1579	Still images	Varying
UIUC2	Badminton	No	Motion type Shot type Shot moment Bounding boxes	3	2m 31s	YouTube
Volleyball Activity Dataset	Volleyball	No	Player detections Action detections	6	23 min	1920 × 1080, 25fps

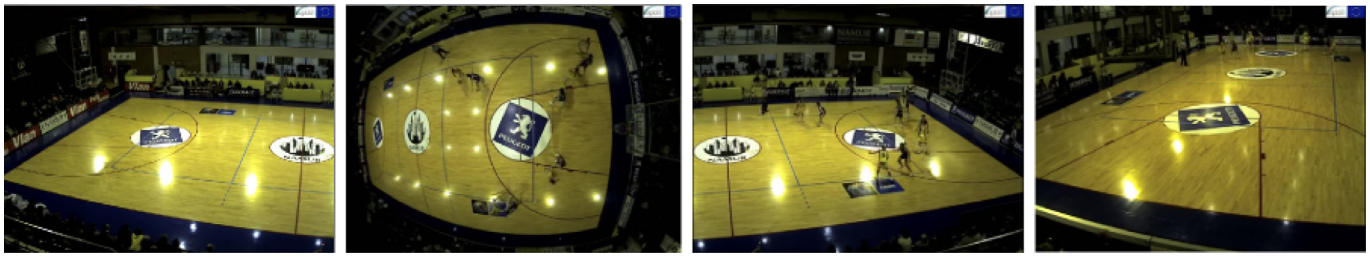


Fig. 6. Example frames from four views from the APIDIS Basketball Dataset (Vleeschouwer et al., 2008).

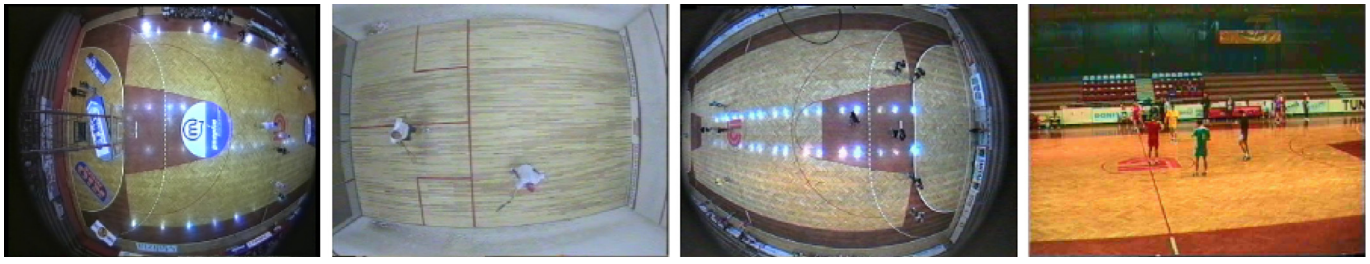


Fig. 7. Example frames from the CVBASE '06 dataset (Pers et al., 2006).

(lob, drop, cross, etc.), stroke outcome (play, let, error), shot type (forehand and backhand).
The third subset consists of two synchronized overhead cameras capturing 5 min of basketball. No annotations are available for the basketball subset. Fig. 7 shows frames from each subset. The dataset is available for download from http://vision.fe.uni-lj.si/cvbase06/downloads.html.

- **ISSIA Soccer Dataset**
The ISSIA Dataset (D'Orazio et al., 2009) contains video from a soccer game captured with 6 cameras, three on each side of the

Fig. 8. Example frames from three views from the ISSIA Dataset (D'Orazio et al., 2009).

Fig. 9. Example images from Leeds Sports Pose Dataset (Johnson and Everingham, 2010).

field. Example frames from one side of the field are shown in Fig. 8. Positions of players, referees, and ball is manually annotated in each from of each camera. Calibration images and measurements are available for calibrating each camera to a common world coordinate system.
The dataset is available for download from http://pspagnolo.jimdo.com/download/.

- **Leeds Sports Pose Dataset**
  This dataset (Johnson and Everingham, 2010) contains 2000 pose-annotated images of mostly sports people gathered from Flickr with the following tags: Athletics, badminton, baseball, gymnastics, parkour, soccer, tennis and volleyball. Each image has been annotated with 14 joint locations for human pose estimation. Example images are presented in Fig. 9.
  The dataset is available for download from http://www.comp.leeds.ac.uk/mat4saj/lsp.html.
- **Olympic Sports Dataset**
  The Olympic Sports Dataset (Niebles et al., 2010) shown in Fig. 10 contains videos of athletes practising 16 different sports, all obtained from YouTube. The dataset contains the following 16 activities: High jump, long jump, triple jump, pole vault, discus throw, hammer throw, javelin throw, shot put, basketball lay-up, bowling, tennis serve, platform diving, springboard diving, snatch (weightlifting), clean and jerk (weightlifting), gymnastic vault.
  50 video sequences are collected for each activity. The data is annotated with the class label for each activity.
  The dataset is available for download from http://vision.stanford.edu/Datasets/OlympicSports/.
- **S-HOCK dataset**
  The S-HOCK dataset (Conigliaro et al., 2015) focuses on the spectators during four hockey matches. 5 cameras are used as shown in Fig. 11: 2 wide angle HD cameras for panoramic views, and three high resolution cameras focusing on different parts of the crowd. Three types of annotation are available: Detection (localizing the body and the head), posture, and action annotation. A total of 13.965 frames and 1.950.210 people are annotated.
  The dataset is available for download from http://vips.sci.univr.it/dataset/shock/.
- **Soccer Video and Player Position Dataset**
  The Soccer Video and Player Position Dataset (Pettersen et al., 2014) is captured during three elite soccer matches. The player positions are measured at 20 Hz using the ZXY Sport Tracking system, and the video is captured from the middle of the field using two camera arrays. The player tracking system provides the player coordinates on the field, their speed, acceleration and force together with an ID and timestamp. The camera array covers the entire field, and each camera can be used individually or as a stitched panorama video. Two full matches (each 90 min) are available with video from three single cam-

Fig. 10. Example frames from Olympic Sports Dataset (Niebles et al., 2010).

Fig. 11. Example frames from the five views of the S-HOCK dataset (Conigliaro et al., 2015).

eras and a stitched panorama video. From the third match 40 min panorama video is available; an example frame shown in Fig. 12. For this sequence the ball position is annotated for a subset of frames.

The dataset is available for download from http://home.ifi.uio.no/paalh/dataset/alfheim/.

- **Sports Videos in the Wild**

  Sports Videos in the Wild (SVW) (Safdarnejad et al., 2015) is comprised of 4200 videos captured with smartphones by users of Coachs Eye smartphone app, an app for sports training developed by TechSmith corporation. SVW includes 30 categories of sports and 44 different actions. The dataset is highly unconstrained with amateur players and unprofessional capturing by amateur users of the app. Example frames from the dataset are shown in Fig. 13.

  Each video is annotated with the sport genre. In addition, for 40% of the video, the time span of each action and a bounding box showing the spatial extent of the action at the start and end frame of the action are also specified.

  The dataset is available for download from http://www.cse.msu.edu/~liuxm/sportsVideo/.

- **Sports-1M Dataset**

  The Sports-1M Dataset (Karpathy et al., 2014) contains 1,133,158 video URLs which have been annotated automatically with 487 Sports labels using the YouTube Topics API. The classes are arranged in a manually-curated taxonomy that contains internal nodes such as Aquatic Sports, Team Sports, Winter Sports, Ball Sports, Combat Sports, Sports with Animals, and generally becomes fine-grained by the leaf level. For example, the dataset contains 6 different types of bowling, 7 different types of Amer-

Fig. 12. Example frame from the panoramic view of Soccer Video and Player Position dataset (Pettersen et al., 2014).

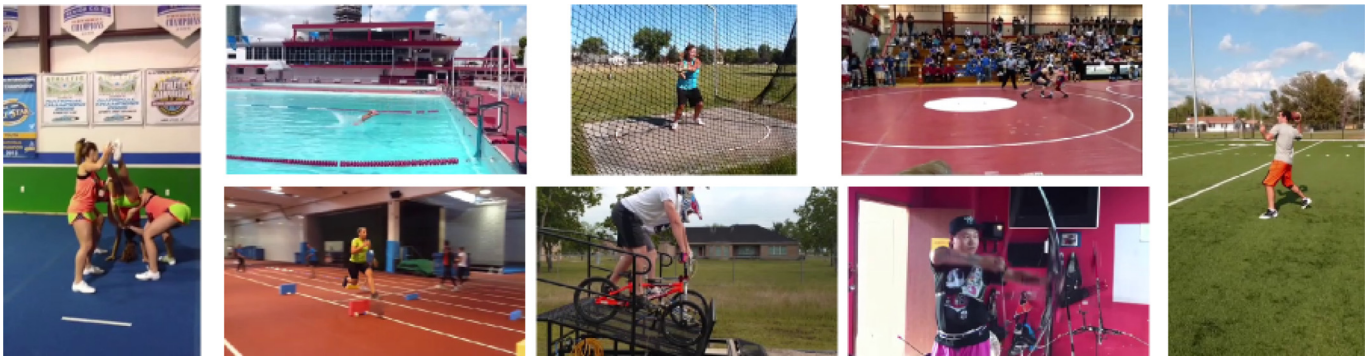


Fig. 13. Example frames from Sports Videos in the Wild dataset (Safdarnejad et al., 2015).

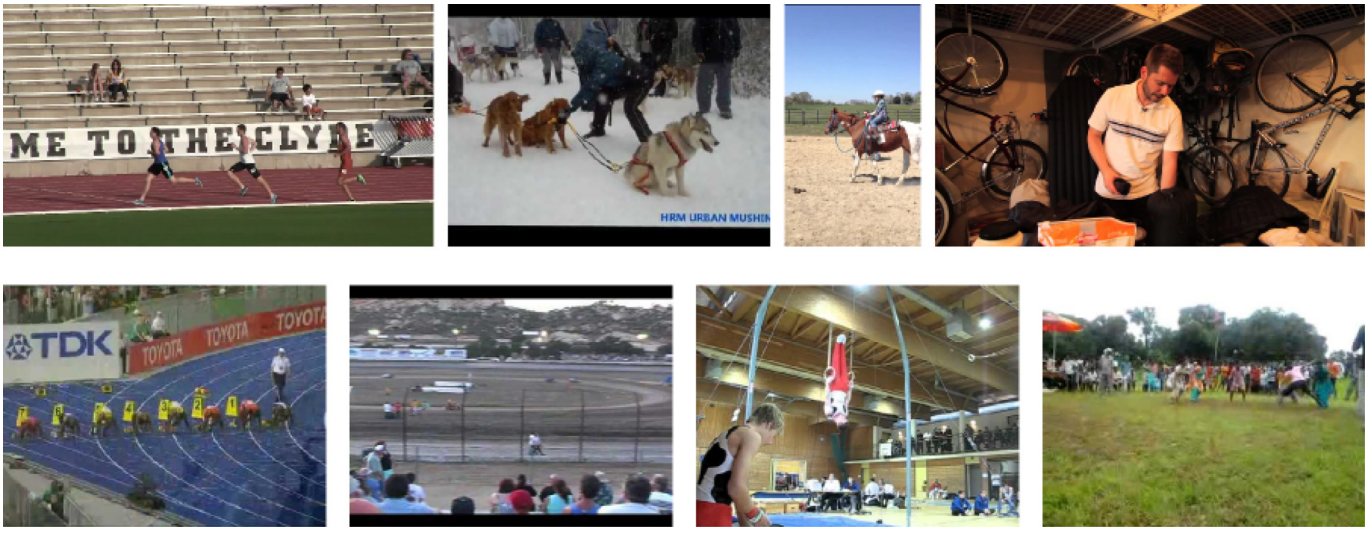


Fig. 14. Example frames from Sports-1M dataset (Karpathy et al., 2014).

ican football and 23 types of billiards. Example frames from the dataset are presented in Fig. 14. There are 1000–3000 videos per class and approximately 5% of the videos are annotated with more than one class. Available in a JSON file is information on each sequence; Duration, resolution and label.

The dataset is available for download from http://cs.stanford.edu/people/karpathy/deepvideo/.

- **UCF Sports Action Dataset**
  UCF Sports Action Dataset (Rodriguez et al., 2008; Soomro and Zamir, 2014) consists of a set of actions collected from vari-

Fig. 15. Example frames from UCF Sports Action dataset (Rodriguez et al., 2008; Soomro and Zamir, 2014).

Fig. 16. Example images from UIUC Sports Event dataset (Li and Fei-Fei, 2007).

ous sports which are typically featured on broadcast television channels such as the BBC and ESPN. The video sequences were obtained from a wide range of stock footage websites including BBC Motion gallery and GettyImages. The dataset contains the following 10 activities: Diving (14 videos), golf swing (18 videos), kicking (20 videos), lifting (6 videos), riding horse (12 videos), running (13 videos), skateboarding (12 videos), swing-bench (20 videos), swing-side (13 videos), walking (22 videos). Example frames are shown in Fig. 15.

The dataset includes a total of 150 sequences with the resolution of 720 x 480 at 10 fps. Available annotations are bounding boxes for action localization and the class label for activity recognition. Furthermore, human gaze annotations from 16 viewers are available.

The dataset is available for download from http://crcv.ucf.edu/data/UCF_Sports_Action.php.

- **UIUC Sports Event Dataset**
  This dataset (Li and Fei-Fei, 2007) contains images from 8 sports event categories shown in Fig. 16: rowing (250 images), badminton (200 images), polo (182 images), bocce (137 images), snowboarding (190 images), croquet (236 images), sailing (190 images), and rock climbing (194 images). The label is provided with each image. Images are divided into easy and medium according to the human subject judgement. Information of the distance of the foreground objects is also provided for each image as close, mid and far.
  The dataset is available for download from http://vision.stanford.edu/lijiali/event_dataset/.
- **UIUC2**
  This dataset (Tran and Sorokin, 2008) contains 3 video sequences obtained from YouTube from the Badminton World Cup 2006: 1 single (3072 frames) and 2 double matches (1648 and 3937 frames). An example frame from each sequence is shown in Fig. 17. 1 sequence is annotated frame-by-frame with five types of motion: Run, walk, hop, jump, and unknown, four types of shot: Forehand, backhand, smash, and unknown, and shot moment: Shot vs. non-shot. Furthermore, foreground masks and bounding boxes of the players are provided for all three sequences.
  The dataset is available for download from http://vision.cs.uiuc.edu/projects/activity/.
- **Volleyball Activity Dataset**
  This dataset (Waltner et al., 2014) contains 6 video sequences captured from games in the professional Austrian Vol-

Fig. 17. Example frames from UIUC2 badminton dataset (Tran and Sorokin, 2008).

Fig. 18. Example frames from Volleyball Activity Dataset (Waltner et al., 2014).

ley League. The videos have a high resolution of 1920 × 1080 pixels at 25 fps, example frames are shown in Fig. 18. The dataset is annotated with bounding boxes and activity labels for each player, using seven possible activities: Serve, reception, setting, attack, block, stand, and defence/move.
The dataset is available for download from http://lrs.icg.tugraz.at/download.php#vb14.

4. Summary

This paper has highlighted some of the current uses for computer vision in sports, and discussed some of the current challenges.

There are now well-established commercial applications using technologies such as multi-camera ball tracking to provide in-depth data for coaching, helping the referee and providing analysis for TV viewers. Vision-based tracking systems for broadcast cameras allow overlays to be placed into the image. Player tracking applications for tactical analysis and coaching tend to rely on semi-automated approaches, with operators helping to initialise and correct vision-based tracking systems. Analysis of the detailed movement of individual competitors still tends to require the use of marker-based motion capture systems, limiting these applications to high-end training scenarios.

Detection and tracking of people or the ball as well as semantic scene understanding are still open research topics when it comes to sports especially for activities where physical interaction is an inherent part of the game. Compared to other detection, tracking and semantic scene understanding applications in computer vision, the world of sports analysis often introduces additional difficulties like rapid and abrupt motion, similar appearance and frequent occlusions. On the other hand, sports applications often have built-in constraints that can make the problems more tractable. These are the physical constraints of the arenas where sports are being performed and the formal (and informal!) rules of each particular (part of a) game. A clear future direction for computer vision research in sports is therefore to enforce the constraints more closely on the different algorithms. Moreover, the combination of the different topics (detection, tracking and semantics) is also a promising avenue to follow. A good example of this is Bialkowski et al. (2014), where it was shown that by first detecting the semantics and then using that as prior could improve the tracking results.

As seen in other computer vision subfields, the availability of public benchmarking datasets is very important. As illustrated in this paper a number of such datasets are already available, but many publications still present results evaluated only on their own datasets. With a continued focus on - and contribution to - publicly available datasets a more unified approach to this field can be expected, helping to drive further advances.

References

Aldridge, D., 2013. SportVU Cameras Shift Focus of What's Possible with NBA Stats. http://www.nba.com/2013/news/features/david_aldridge/11/11/morning-tip-sportvu-cameras-in-arenas-problems-with-nets-qa-with-paul-george/. Accessed 12 Feb 2017.

Anderson, S., 2013. Forsgren helps revolutionize golf on TV with Protracer. Golf WRX.

Atmosukarto, I., Ghanem, B., Saadalla, M., Ahuja, N., 2014. Recognizing team formation in american football. In: Moeslund, T., Thomas, G., Hilton, A. (Eds.), Computer Vision in Sports, chapter 13. Springer.

Badminton, B., 2017. Hawk-eye to Determine "In or Out". BWF Badminton, 04 April, 2014 http://bwfbadminton.com/2014/04/04/hawk-eye-to-determine-in-or-out/. Accessed 26 Feb 2017.

Behendi, S.K., Morgan, S., Fookes, C.B., 2016. Non-invasive performance measurement in combat sports. In: Chung, P., Soltoggio, A., Dawson, W.C., Meng, Q., Pain, M. (Eds.), Proceedings of the 10th International Symposium on Computer Science in Sports (ISCSS). Springer International Publishing, pp. 3–10.

Bialik, C., 2014. The people tracking every touch, pass and tackle in the world cup. Five-Thirty-Eight.

Bialkowski, A., Lucey, P., Carr, P., Sridharan, S., Matthews, I., 2014. Representing team behaviours from noisy data using player role. In: Moeslund, T., Thomas, G., Hilton, A. (Eds.), Computer Vision in Sports, chapter 12. Springer.

Boril, H., Sangwan, A., Hasan, T., Hansen, J.H.L., 2010. Automatic excitement-level detection for sports highlights generation. In: 11th Annual Conference of the International Speech Communication Association (INTERSPEECH).

Broussard, C., 2017. Pro Basketball; N.B.A. will Use Replay to Review Buzzer Shots. The New York Times. July 30, 2002. http://www.nytimes.com/2002/07/30/sports/pro-basketball-nba-will-use-replay-to-review-buzzer-shots.html. Accessed 12 Feb 2017.

Chakraborty, B., Meher, S., 2012. Real-time position estimation and tracking of a basketball. In: IEEE International Conference on Signal Processing, Computing and Control.

Chakraborty, B., Meher, S., 2013. A real-time trajectory-based ball detection-and–tracking framework for basketball video. J. Opt. 42 (2), 156–170.

Chang, Y., Maheswaran, R., Su, J., Kwok, S., Levy, T., Wexler, A., Squire, K., 2014. Quantifying shot quality in the NBA. In: MIT Sloan Sports Analytics Conference.

Chen, F., Delannay, D., Vleeschouwer, C.D., 2011. An autonomous framework to produce and distribute personalized team-sport video summaries: a basketball case study. IEEE Trans. Multim. 13 (6), 1381–1394.

Chen, H., Tien, M., Chen, Y., Tsai, W., Lee, S., 2009. Physics-based ball tracking and 3d trajectory reconstruction with applications to shooting location estimation in basketball video. J. Vis. Commun. Image Represent. 20 (3), 204–216.

ChyronHego, 2017a. 5 Reasons You Need TRACAB Player Tracking. http://chyronhego.com/sports-data/tracab. Accessed 12th Feb 2017.

ChyronHego, 2017b. Product Information Sheet TRACAB Optical Tracking. http://chyronhego.com/sports-data/tracab. Accessed 12th Feb 2017.

Cole, B., 2014. Making sense of the video tracking systems. Beyond the Box Score.

Conigliaro, D., Rota, P., Setti, F., Bassetti, C., Conci, N., Sebe, N., Cristani, M., 2015. The s-hock dataset: analyzing crowds at the stadium. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Dawes, R., Chandaria, J., Thomas, G., 2009. Image-based camera tracking for athletics. In: Proceedings of the IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB 2009) Available as BBC R&D White Paper 181.

D'Orazio, T., Leo, M., Mosca, N., Spagnolo, P., Mazzeo, P.L., 2009. A semi-automatic system for ground truth generation of soccer video sequences. In: IEEE International Conference on Advanced Video and Signal Surveillance (AVSS).

Ekin, A., Tekalp, A.M., Mehrotra, R., 2003. Automatic soccer video analysis and summarization. IEEE Trans. Image Process. 12 (7), 796–807.

EPFL, 2016. Tracking Multiple People in a Multi-Camera Environment. http://cvlab.epfl.ch/research/body/surv. Accessed 7th July 2016.

Ericsson, 2016. The Piero™Sports Graphics System. http://www.ericsson.com/broadcastandmedia/what-we-do/piero. Accessed 2 May 2016.

Stack Exchange, S., 2017. Why is FIFA Against Adding Instant Replay to the Game?. http://sports.stackexchange.com/questions/179/why-is-fifa-against-adding-instant-replay-to-the-game. Accessed 12 Feb 2017.

Fast, M., 2010. What the heck is pitchf/x? The Hardball Times Baseball Annual ACTA Publications.

Faulkner, H., Dick, A., 2015. AFL player detection and tracking. In: International Conference on Digital Image Computing: Techniques and Applications (DICTA).

Fetters, A., 2012. How Instant Replays Changed Professional Tennis. The Atlantic, Sep 7, 2012. https://www.theatlantic.com/entertainment/archive/2012/09/how-instant-replays-changed-professional-tennis/262060/. Accessed 12 Feb 2017.

FIFA, 2016a. FIFA Certified GLT Installations. http://quality.fifa.com/en/Goal-Line-Technology/FIFA-certified-GLT-installations. Accessed 10 May 2016.

FIFA, 2014. FIFA Quality Programme for Goal-Line Technology - Testing Manual 2014. http://quality.fifa.com/en/Goal-Line-Technology/Get-to-know-goal-line-technology/See-how-it-is-tested/. Accessed 12 Feb 2017.

FIFA, 2016c. Goal-line Technology Set Up Ahead of FIFA World Cup. http://www.fifa.com/worldcup/news/y=2014/m=4/news=goal-line-technology-set-ahead-fifa-world-cup-2311481.html. Accessed 7th July 2016.

Figueroa, P.J., Leite, N.J., Barros, R.M.L., 2006. Tracking soccer players aiming their kinematical motion analysis. Comput. Vision Image Understanding 101, 122–135.

Fu, T.S., Chen, H.T., Chou, C.L., Tsai, W.J., Lee, S.Y., 2011. Screen-strategy analysis in broadcast basketball video using player tracking. In: IEEE Visual Communications and Image Processing (VCIP), pp. 1–4.

Gade, R., Moeslund, T.B., 2014. Classification of sports types using thermal imagery. In: Moeslund, T.B., Thomas, G., Hilton, A. (Eds.), Computer Vision in Sports. Springer. chapter 10.

Gade, R., rgensen, A.J., Moeslund, T.B., 2012. Occupancy analysis of sports arenas using thermal imaging. In: Proceedings of the International Conference on Computer Vision Theory and Applications.

Gerke, S., Müller, K., Schäfer, R., 2015. Soccer jersey number recognition using convolutional neural networks. In: IEEE International Conference on Computer Vision Workshop (ICCVW).

Gilligan, J., 2014. How GPS Technology is Changing Rugby. SportTechie, May 19, 2014 http://www.sporttechie.com/2014/05/19/uncategorized/how-gps-technology-is-changing-rugby/. Accessed 12 Feb 2017.

Grau, O., Price, M., Thomas, G., 2001. Use of 3-D techniques for virtual production. In: SPIE Conference on Videometrics and Optical Methods for 3D Shape Measurement.

Haimovitch, O., 2015. System and Method of Limiting Processing by a 3d Reconstruction System of an Environment in a 3d Reconstruction of an Event Occurring in an Event Space. US Patent application US2016189421.

Halbinger, J., Metzler, J., 2015. Video-based soccer ball detection in difficult situations. Communications in Computer and Information Science. Sports Science Research and Technology Support: International Congress, 464.

Hall, A., 2014. ESPN emerging technologies enhancement of golf visuals displayed at open championship. ESPN Front Row.

Hanjalic, A., 2005. Adaptive extraction of highlights from a sport video based on excitement modeling. IEEE Trans. Multimedia. 7 (6), 1114–1122.

IIS, F., 2016. Goalref™Goal Detection System. http://www.iis.fraunhofer.de/en/ff/kom/proj/goalref.html. Accessed 7th July 2016.

Innovations, H.-E., 2017a. Hawk-Eye Electronic Line Calling. http://www.hawkeyeinnovations.co.uk/products/ball-tracking/electronic-line-calling. Accessed 12 Feb 2017.

Innovations, H.-E., 2017b. Hawk-Eye Goal Line Technology. http://www.hawkeyeinnovations.co.uk/products/ball-tracking/goal-line-technology. Accessed 12 Feb 2017.

Innovations, H.-E., 2017c. Hawk-Eye in Cricket. http://www.hawkeyeinnovations.co.uk./sports/cricket. Accessed 12 Feb 2017.

Innovations, H.-E., 2017d. Hawk-Eye Tennis System. http://www.hawkeyeinnovations.co.uk/sports/tennis. Accessed 12 Feb 2017.

Innovations, H.-E., 2017e. Paul Hawkins Response to ESPN CricInfo Blog "Why Ball-Tracking can't be Trusted". http://www.hawkeyeinnovations.co.uk./sports/cricket. Accessed 12 Feb 2017.

Ivankovic, Z., Markoski, B., Ivkovic, M., Radosav, D., Pecev, P., 2012. Adaboost in basketball player identification. IEEE 13th International Symposium on Computational Intelligence and Informatics.

Johnson, S., Everingham, M., 2010. Clustered pose and nonlinear appearance models for human pose estimation. In: Proceedings of the 21st British Machine Vision Conference (BMVC).

Kapela, R., McGuinness, K., Swietlicka, A., O'Connor, N.E., 2014. Real-time event detection in field sport videos. In: Moeslund, T., Thomas, G., Hilton, A. (Eds.), Computer Vision in Sports, chapter 14. Springer.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L., 2014. Large-scale video classification with convolutional neural networks. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR).

Kasuya, N., Kitahara, I., Kameda, Y., Ohta, Y., 2008. Robust trajectory estimation of soccer players by using two cameras. In: 19th International Conference on Pattern Recognition.

Kolekar, M.H., Sengupta, S., 2015. Bayesian network-based customized highlight generation for broadcast soccer videos. IEEE Trans. Broadcast. 61 (2), 195–209.

Krishna Mohan, C., Yegnanarayana, B., 2010. Classification of sport videos using edge-based features and autoassociative neural network models. Signal Image Video Process. 4, 61–73.

Kristan, M., Pers, J., Perse, M., Kovacic, S., 2009. Closed-world tracking of multiple interacting targets for indoor-sports applications. Comput. Vision Image Understand. 113 (5), 598–611.

Lee, J.Y., Hoff, W., 2007. Activity identification utilizing data mining techniques. IEEE Workshop on Motion and Video Computing (WMVC) doi:10.1109/WMVC.2007.4.

Leo, M., Mosca, N., Spagnolo, P., Mazzeo, P., D'Orazio, T., Distante, A., 2008. Real-time multiview analysis of soccer matches for understanding interactions between ball and players. In: Conference on Image and Video Retrieval.

Leung, D., 2017. NBA Teams Banned from using Wearables Data in Contract Negotiations, Player Transactions. SportTechie, January 31, 2017 http://www.sporttechie.com/2017/01/31/sports/nba/nba-teams-banned-using-wearables-data-contract-negotiations-player-transactions/. Accessed 12 Feb 2017.

Li, L., Fei-Fei, L., 2007. What, where and who? Classifying event by scene and object recognition. In: IEEE International Conference in Computer Vision (ICCV).

Liu, J., Carr, P., 2014. Detecting and tracking sports players with random forests and context-conditioned motion models. In: Moeslund, T.B., Thomas, G., Hilton, A. (Eds.), Computer Vision in Sports. Springer. chapter 6.

Lowe, Z., 2013. Lights, cameras, revolution. Grantland.

Lu, W., Ting, J., Murphy, K.P., Little, J.J., 2011. Identifying players in broadcast sports videos using conditional random fields. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR).

Maksai, A., Wang, X., Fua, P., 2016. What players do with the ball: A physically constrained interaction modeling. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR).

Manafifard, M., Ebadi, H., Moghaddam, H.A., 2016. A survey on player tracking in soccer videos. Submitted for Computer Vision and Image Understanding, Special Issue on Computer Vision in Sports.

McQueen, A., Wiens, J., Guttag, J., 2014. Automatically recognizing on-ball screens. In: MIT Sloan Sports Analytics Conference.

McSurley, K., Rybarczyk, G., 2011. An Introduction to FIELDf/x. The Hardball Times Baseball Annual ACTA Publications.

Mo-Sys, 2016. Red Bull Air Race with Gyrotracker. http://www.mo-sys.com/news/red-bull-air-race-gyrotracker. Accessed 15 May 2016.

Mutchima, P., Sanguansat, P., 2012. TF-RNF: a novel term weighting scheme for sports video classification. In: IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC) doi:10.1109/ICSPCC.2012.6335651.

Nguyen, N., Yoshitaka, A., 2014. Soccer video summarization based on cinematography and motion analysis. In: IEEE 16th International Workshop on Multimedia Signal Processing (MMSP), pp. 1–6.

Niebles, J.C., Chen, C., Fei-Fei, L., 2010. Modeling temporal structure of decomposable motion segments for activity classification. In: 11th European Conference on Computer Vision (ECCV).

Orad, 2016. Trackvision First Down Line System. http://www.orad.tv/product/sports/fdl. Accessed 11 May 2016.

Owens, N., Harris, C., Stennett, C., 2003. Hawk-eye tennis system. In: International Conference on Visual Information Engineering (VIE2003), pp. 182–185.

Pers, J., Bon, M., Vuckovic, G., 2006. Cvbase '06 Dataset. http://vision.fe.uni-lj.si/cvbase06/downloads.html.

Perse, M., Kristan, M., Kovacic, S., Vuckovic, G., Pers, J., 2009. A trajectory-based analysis of coordinated team activity in a basketball game. Comput. Vision Image Understand. 113 (5), 612–621. Computer Vision Based Analysis in Sport Environments.

Petrovic, K., 2017. Intel 360-degree Replay Technology Brings Basketball Fans into the Future. https://iq.intel.com/360-degree-replay-technology-brings-fans-into-the-future-of-sports/. Accessed 12 Feb 2017.

Pettersen, S.A., Johansen, D., Johansen, H., Berg-Johansen, V., Gaddam, V.R., Mortensen, A., Langseth, R., Griwodz, C., Stensland, H.K., Halvorsen, P., 2014. Soccer video and player position dataset. In: Proceedings of the International Conference on Multimedia Systems (MMSys).

Popa, T., Germann, M., Ziegler, R., Keiser, R., Gross, M., 2014. Geometry reconstruction of players for novel-view synthesis of sports broadcasts. In: Moeslund, T.B., Thomas, G., Hilton, A. (Eds.), Computer Vision in Sports. Springer. chapter 7.

Prandoni, P., 2003. Automated Stroboscoping of Video Sequences. European Patent Specification EP 1287.

Rao, M.U., Pati, U.C., 2015. A novel algorithm for detection of soccer ball and player. In: International Conference on Communications and Signal Processing.

R&D, B., 2017. Piero Sports Graphics System. http://www.bbc.co.uk/rd/projects/piero. Accessed 12 Feb 2017.

Ren, J., Orwell, J., Jones, G.A., Xu, M., 2009. Tracking the soccer ball using multiple fixed cameras. Comput. Vision Image Understand. 113 (5), 633–642.

Reno, V., Mosca, N., Nitti, M., D'Orazio, T., Campagnoli, D., Prati, A., Stella, E., 2015. Tennis player segmentation for semantic behavior analysis. In: IEEE International Conference on Computer Vision Workshop (ICCVW).

Reusens, M., Vetterli, M., Ayer, S., Bergnozoli, V., 2007. Coordination and Combination of Video Sequences with Spatial and Temporal Normalization. European Patent Specification EP 4, 2007.

Rodriguez, M.D., Ahmed, J., Shah, M., 2008. Action mach: a spatio-temporal maximum average correlation height filter for action recognition. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR).

Safdarnejad, S.M., Liu, X., Udpa, L., Andrus, B., Wood, J., Craven, D., 2015. Sports videos in the wild (svw): a video dataset for sports analysis. In: 11th IEEE Int. Conf. Automatic Face and Gesture Recognition.

Soomro, K., Khokhar, S., Shah, M., 2015. Tracking when the camera looks away. In: IEEE International Conference on Computer Vision Workshop (ICCVW).

Soomro, K., Zamir, A.R., 2014. Action recognition in realistic sports videos. In: Moeslund, T.B., Thomas, G., Hilton, A. (Eds.), Computer Vision in Sports. Springer. chapter 9.

de Sousa, S.F., de A. Arajo, A., Menotti, D., 2011. An overview of automatic event detection in soccer matches. In: IEEE Workshop on Applications of Computer Vision (WACV), pp. 31–38.

Sportvision, 2016a. https://www.sportvision.com/football/1st-ten-system. Accessed 11 May 2016.

Sportvision, 2016b. Liveline. http://www.sportvision.com/sailing/liveline. Accessed 15 May 2016.

STATS, 2017. STATS SportVU ®: Player Tracking and Predictive Analytics. https://www.stats.com/publications/stats-sportvu-player-tracking-advanced-analytics/. Accessed 12 Feb 2017.

Storey, R., 1984. TELETRACK - A Special Effect. BBC Research Department Report 1984-10, Available as BBC R&D White Paper 033.

Takahashi, D., 2017. Those Super Bowl Instant Replays You'll Love Run off Intels Tech. http://venturebeat.com/2016/02/06/super-bowls-eyevision-instant-replays-from-any-angle-are-powered-by-intel/. Accessed 12 Feb 2017.

Tamaki, S., Saito, H., 2014. Plane approximation-based approach for 3d reconstruction of ball trajectory for performance analysis in table tennis. In: Moeslund, T.B., Thomas, G., Hilton, A. (Eds.), Computer Vision in Sports, chapter 3. Springer.

Tamir, M., Oz, G., 2008. Real-Time Objects Tracking and Motion Capture in Sports Events. US Patent App. 11/909,080.

Technologies, R., 2017. Freed^{TM}Free Dimensional Video. http://replay-technologies.com/. Accessed 12 Feb 2017.

Thomas, G., 2007. Real-time camera tracking using sports pitch markings. J. Real Time Image Processing 2, 2–3.

Tran, D., Sorokin, A., 2008. Human Activity Recognition with Metric Learning. In: European Conference on Computer Vision (ECCV).

UEFA, 2016. Uefa Euro 2016 to Use Hawk-Eye for Goal-Line Technology. http://www.uefa.org/mediaservices/mediareleases/newsid=2354960.html. Accessed 7th July 2016.

Mellon University, C., 2017. Carnegie Mellon goes to the Super Bowl. https://www.ri.cmu.edu/events/sb35/tksuperbowl.html. Accessed 12 Feb 2017.

Vizrt, 2016. http://www.vizrt.com/products/viz_libero/. Accessed 11 May 2016.

Vleeschouwer, C.D., Chen, F., Delannay, D., Parisot, C., Chaudy, C., Martrou, E., Cavallaro, A., 2008. Distributed video acquisition and annotation for sport-event summarization. NEM Summit: Towards Future Media Internet.

Waltner, G., Mauthner, T., Bischof, H., 2014. Indoor activity detection and recognition for automated sport games analysis. In: Austrian Conference on Pattern Recognition (AAPR/OAGM).

Wang, X., Turetken, E., Fleuret, F., Fua, P., 2014. Tracking interacting objects optimally using integer programming. In: European Conference on Computer Vision (ECCV).

Wang, X., Turetken, E., Fleuret, F., Fua, P., 2016. Tracking interacting objects using intertwined flows. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Wei, X., Sha, L., Lucey, P., Carr, P., Sridharan, S., Matthews, I., 2015. Predicting ball ownership in basketball from a monocular view using only player trajectories. In: 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), pp. 780–787.

Wilson, S., Mohan, C.K., Murthy, K.S., 2014. Event-based sports videos classification using HMM framework. In: Moeslund, T.B., Thomas, G., Hilton, A. (Eds.), Computer Vision in Sports, chapter 11. Springer.

Wolverton, T., 2017. Big Data Meets Big-Time Basketball. http://www.mercurynews.com/2014/05/17/big-data-meets-big-time-basketball/. Accessed 12 Feb 2017.

Wu, L., Meng, X., Liu, X., Chen, S., 2006. A new method of object segmentation in the basketball videos. In: 18th International Conference on Pattern Recognition (ICPR).

Xiao, J., Stolkin, R., Leonardis, A., 2014. Multi-target tracking in team-sports videos via multi-level context-conditioned latent behaviour models. In: British Machine Vision Conference (BMVC).

Xu, M., Orwell, J., Jones, G., 2004. Tracking football players with multiple cameras. In: International Conference on Image Processing.

Yan, F., Christmas, W., Kittler, J., 2014. Ball tracking for tennis video annotation. In: Moeslund, T.B., Thomas, G., Hilton, A. (Eds.), Computer Vision in Sports. Springer, Advances in Computer Vision and Pattern Recognition. chapter 2.

Ye, Q., Huang, Q., Jiang, S., Liu, Y., Gao, W., 2005. Jersey number detection in sports video for athlete identification. Visual Commun. Image Process. Proc. SPIE 5960.

Yuan, Y., Wan, C., 2004. The application of edge feature in automatic sports genre classification. In: IEEE Conference on Cybernetics and Intelligent Systems doi:10.1109/ICCIS.2004.1460749.

Zawbaa, H.M., El-Bendary, N., Hassanien, A.E., Abraham, A., 2011. Svm-based soccer video summarization system. In: Third World Congress on Nature and Biologically Inspired Computing (NaBIC), pp. 7–11.

Zhang, T., Ghanem, B., Ahuja, N., 2012. Robust multi-object tracking via cross-domain contextual information for sports video analysis. In: IEEE International Conference on Acoustics, Speech and Signal Processing.