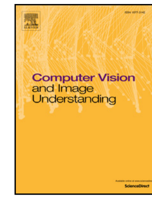
ARTICLE IN PRESS
JID: YCVIU [m5G;April 26, 2017;21:31]

Computer Vision and Image Understanding 000 (2017) 1–16

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 xxx

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/© 2017 Published by Elsevier Inc.

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

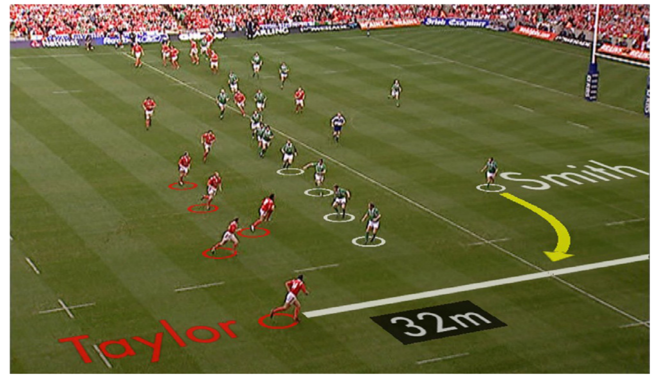


Fig. 2. Example of virtual graphics overlaid on a rugby pitch [picture courtesy of Ericsson].

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-

Please cite this article as: G. Thomas et al., Computer vision for sports: Current applications and research topics, Computer Vision and Image Understanding (2017), http://dx.doi.org/10.1016/j.cviu.2017.04.011