Computer Vision and Image Understanding 119 (2014) 102–115

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

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

Take your eyes off the ball: Improving ball-tracking by focusing on team play ☆

Xinchao Wang *, Vitaly Ablavsky, Horesh Ben Shitrit, Pascal Fua

Computer Vision Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), EPFL/IC/ISIM/CVLab, Station 14, CH-1015 Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 20 March 2013
Accepted 30 November 2013
Available online 12 December 2013

Keywords:
Tracking
Video analysis and event recognition
Occlusion reasoning
Behavior analysis
Sports video
Tracking the invisible

ABSTRACT

Accurate video-based ball tracking in team sports is important for automated game analysis, and has proven very difficult because the ball is often occluded by the players. In this paper, we propose a novel approach to addressing this issue by formulating the tracking in terms of deciding which player, if any, is in possession of the ball at any given time. This is very different from standard approaches that first attempt to track the ball and only then to assign possession. We will show that our method substantially increases performance when applied to long basketball and soccer sequences.

1. Introduction

Accurate ball tracking in sports is of tremendous importance to athletes, referees, coaches, and fans. However, the size and speed of the ball and prolonged occlusions make ball-tracking challenging even for trained human observers. Solutions for basketball or soccer have not gained acceptance, and a solution for hockey puck tracking [1] developed for a television broadcast network enjoyed only a brief lifespan.

Thus, a Computer Vision-based ball-tracking solution remains valuable and beyond the state of the art for most sports. Tennis is a rare exception because the ball is rarely occluded and its color is very different from that of the background, which makes tracking comparatively easy. It is far more difficult in team sports because the ball is often hidden by the players and follows an unpredictable trajectory as it is passed or taken from one player to another. Furthermore, the amount of image evidence for the ball, even when it is visible, is dwarfed by that for the players in its vicinity, as shown in Fig. 1. Frame-to-frame tracking is extremely unreliable in such cases. For example, even a state-of-the-art algorithm [2] that has been shown to be superior to many other state-of-the-art trackers for single object tracking and whose code is publicly available never tracks the ball for more than five consecutive frames in the sequences we present in this paper. Modern tracking-by-detection approaches that re-detect the object as often as necessary and aggregate results over several frames increase robustness but can still easily fail if the object is hard to detect in individual frames.

The contribution of this paper is to turn the common approach of first tracking the ball and then deciding which player is in possession of it on its head. In our approach, we first track the players and decide possession. We then use this as a means to achieve reliable ball tracking, which lets us turn unreliable image-based information into dependable trajectories. This is effective because the trajectory of the ball is intimately linked to that of the players who pass it to each other or steal it from one another. Therefore, exploiting this correlation allows for disambiguation and improved performance.

To this end we define a novel state space explicitly accounts for ball possession, a Conditional Random Field (CRF) [3] model that depends on the ball detections and the players' trajectories, and a practical approach to learning the model parameters from training data. More specifically, we start from video sequences from several synchronized cameras from which both the players' trajectories can be reliably extracted using publicly available software [4,5] and the ball ballistic trajectory can be also be extracted in consecutive frames when it is clearly visible. Between such frames, when individual players are in possession of the ball, its approximate position can be assumed to be their position up to the reach of their arms or legs. When the ball is passed, we take its path in the horizontal plane to be a straight line from one player to the next. As shown in Fig. 1, given perfect knowledge of ball possession as denoted by the "Oracle", this simple approach yields a relatively

☆ This paper has been recommended for acceptance by Nikos Paragios.

* Corresponding author.
E-mail addresses: xinchao.wang@epfl.ch (X. Wang), vitaly.ablavsky@epfl.ch (V. Ablavsky), horesh.benshitrit@epfl.ch (H.B. Shitrit), pascal.fua@epfl.ch (P. Fua).

1077-3142/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cviu.2013.11.010

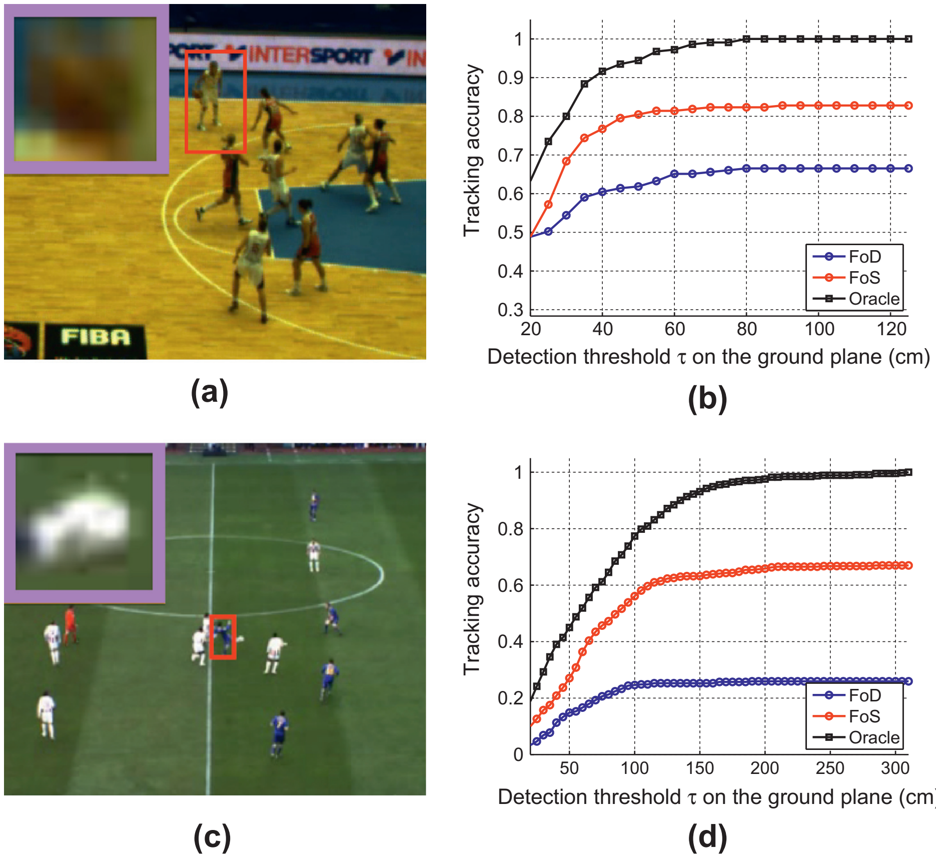

Fig. 1. Tracking the ball in team sports, such as basketball (a) and soccer (c) is hard due to the ball's small size, barely 12 × 12 pixels in the videos we work with as shown in upper left corner insets. Furthermore, it is subject to prolonged occlusions even when multiple views are available. Our key contribution is to overcome these difficulties by exploiting the contextual relation between the players and the ball, in particular with respect to *ball possession*. To illustrate this, we plot in (b) and (d) the accuracy we would obtain based on the sole knowledge of ball possession if it were given to us by an oracle, in this case a person looking at the videos. In other words, the "Oracle" tracker is the one with correct ball possessions, whose performance is the upper-bound of any tracker that is solely based on contextual information. We also plot the accuracy results of both a state-of-the-art approach that only looks at the ball and our attempt to establish ball possession without human input. The latter (red FoS curve) decisively outperforms the former (blue FoD curve) and approaches human performance levels (black Oracle curve). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

accurate estimate of the ball location and obtaining this knowledge therefore is what must be addressed. We formulate this problem in terms of assigning possession to a player, or to no-one when the ball is being passed, by minimizing a global objective function that takes into account players' positions and image information when the ball is visible.

We will show that this approach produces very reliable possession assignments and a final accuracy of the ball trajectory that is close to the best that can be expected over multiple attack phases and timeouts. This is in stark contrast to most other published approaches that have only been demonstrated on short sequences.

2. Related work

This section distinguishes two broad classes of approaches, those whose sole focus is on ball-tracking, and those that exploit spatio-temporal context to track a hard-to-discern target.

2.1. Tracking only the ball

We first discuss a handful of ball-tracking scenarios where the problem is considered solved, and off-the-shelf commercial solutions exist. Then we discuss the more broad class of problems, which are still beyond the reach of the current state of the art.

2.1.1. Ball tracking in the commercial world

Ball tracking has received considerable attention over the years. In some cases such as tennis, reliable commercial software is now available [6]. While the high velocity of the ball poses an engineering challenge, it is brightly colored and rarely occluded. Even then, achieving the required level of reliability and accuracy requires ten high-speed cameras looking at the scene from different angles.

The problem becomes substantially harder in team sports. For example in soccer, the ball can be surrounded by multiple players, who create occlusions in many views. This is probably why the soccer software sold by the company that developed the tennis ball tracking system discussed above [6] only aims at tracking the ball as it is being shot towards the goal, and therefore unoccluded.

In basketball, the problem is even harder because multiple players compete for the possession of the ball, generating even more occlusions. Since the basketball may be guided by hand, its trajectory is also more unpredictable. Lastly, the ball can be of similar color as the players' jerseys, adding yet another level of complexity. Even though some companies advertise the capabilities of post-game basketball trajectory analysis the amount of automation is unclear [7].

2.1.2. Ball tracking in the research world

Ball tracking has also been pursued in the academic world [8–10] before the advent of the commercial systems. For sports