

Proceedings

Using Artificial Vision Techniques for Individual Player Tracking in Sport Events [†]

Roberto López Castro * and Diego Andrade Canosa

Departamento de Ingeniería de Computadores, Universidade da Coruña, Campus de Elviña, 15071 A Coruña, Spain; diego.andrade@udc.es

* Correspondence: roberto.lopez.castro@udc.es; Tel.: +34-981-167-000

† Presented at the 2nd XoveTIC Congress, A Coruña, Spain, 5–6 September 2019.

Published: 31 July 2019



Abstract: We introduce a hybrid approach that can track an individual football player in a video sequence. This solution achieves a good balance between speed and accuracy, combining traditional object tracking techniques with Deep Neural Networks (DNN). While traditional techniques lack accuracy, the main shortcoming of DNN is performance. Both types of techniques complement to each other to provide an accurate and fast object tracking approach that does not require human intervention. The accuracy of our solution has been validated using the SoccerNet Dataset against hand annotated video sequences. For the tracking of 4 different players of 2 different teams our approach has achieved an Area Under Curve (AUC) of 0.66, in terms of accuracy, and a frame rate of 91.75 FPS, in terms of performance, running on a Nvidia GTX 1080Ti GPU.

Keywords: artificial vision; object tracking; object detection; machine learning; deep learning; real time

1. Introduction

The tracking of individual players in sport events is really interesting for coaches, personal trainers, fans and media. One of the best ways to do it automatically is using computer vision [1]. However, the sport case is particularly challenging due to several factors: some players have a very similar aspect, the jersey number is not always visible, the video codification algorithms frequently generate blurry video segments, the player is often partially or totally occluded, etc.

Object tracking algorithms can be classified in two main classes:

- Traditional algorithms based on mathematical and machine learning principles usually suffer lack of accuracy, caused by: the accumulation of tracking errors, which makes the bounding box (area which the algorithm uses to delimit the object) to lose progressively the tracked object, and partial or total occlusions of the tracked individual with others. Additionally, it needs a human operator that makes the initial identification and selection of the tracked individual. A good example of these algorithms are Discriminative Correlation Filters (DCF) [2].
- Deep Neural Networks that can track an object by detecting it in each frame. Specifically, Convolutional Neural Networks (CNNs) [3] are used to solve this problem. A properly trained network can achieve a very good accuracy but at the cost of high computational cost, which makes them often unusable to process high definition video sequences at real-time.

The solution proposed in this work combines two CNNs with one DCF algorithm to perform a fast and accurate tracking of a football player in a video sequence. Besides, the initial position of the individual to be tracked does not have to be selected by a human operator. The solution is fast enough to process video sequences of 60 fps (or more) at real-time, and it is sufficiently accurate to recover

from temporary tracking errors, and to support camera movements and switches from one camera to another.

2. Hybrid Solution

The two CNNs models used in our hybrid solution are Faster-RCNN [4] and SSD [5]. Faster-RCNN is a highly accurate detector but which needs near 45 ms to process a single frame of the video sequence, this means that it can only process 22 fps. On the other hand, SSD is less accurate but has an affordable performance. These two networks are combined in the following manner: Faster-RCNN is executed on the whole frame, but only processes one of every λ frames. In the $\lambda - 1$ frames in between, SSD is applied on a sub-frame cropped around the area where Faster-RCNN detected the tracked individual.

This combination of both CNNs increases performance, but loses accuracy with respect to using Faster-RCNN for every frame. To increase the accuracy of our hybrid approach we add a DCF algorithm, specifically KCF (Kernalized Correlation Filter) [6], to the workflow. This traditional algorithm is good at tracking a previously selected object for some time, but it suffers the aforementioned accuracy problems of this type of algorithms. In our proposal, the two CNNs can play the role of a human operator which is constantly informing KCF of the position of the tracked object. Figure 1 shows the execution diagram of our approach. Faster-RCNN is executed in one of every λ frame playing the role of the guide of the other two algorithms (KCF and SSD). In the remaining $\lambda - 1$ iterations, these other two algorithms collaborate to track the object, SSD constantly correcting, if necessary, the possible tracking errors introduced by KCF.

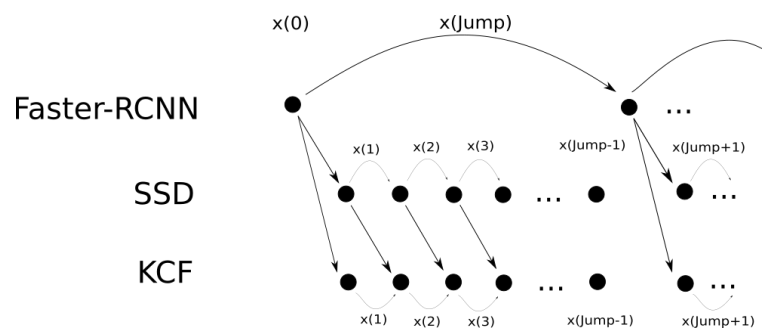


Figure 1. Execution diagram.

3. Results

Our approach has been trained for tracking 4 different players of 2 different teams, using the SoccerNet Dataset [7]. Table 1 shows the average accuracy and performance results obtained when running the algorithm on a NVidia GTX 1080Ti GPU.

Table 1. Average hybrid algorithm performance.

	Avg. Accy	Avg. Fps	Avg. AUC	Lost Frames
Player 1	0.620	91.75	0.610	2
Player 2	0.653	84.98	0.651	0
Player 3	0.650	86.65	0.660	0
Player 4	0.600	87.36	0.600	0
TOTAL AVG.	0.6308	87.685	0.6302	0.5

The performance results show that the approach can process around 87 FPS on average. Regarding the accuracy, the average AUC is 0.6302, a similar value to the one obtained by state-of-the-art algorithms on generic datasets [8].

References

1. Manafifard, M.; Ebadi, H.; Moghaddam, H.A. A survey on player tracking in soccer videos. *Comput. Vis. Image Underst.* **2017**, *159*, 19–46.
2. Lukezic, A.; Vojir, T.; Cehovin Zajc, L.; Matas, J.; Kristan, M. Discriminative correlation filter with channel and spatial reliability. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* **2017**, 6309–6318.
3. Géron, A. *Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2017.
4. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, 91–99.
5. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In *European Conference on Computer Vision*; Springer: Berlin, Germany, 2016; pp. 21–37.
6. Henriques, J.F.; Caseiro, R.; Martins, P.; Batista, J. High-speed tracking with kernelized correlation filters, *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 583–596.
7. Giancola, S.; Amine, M.; Dghaily, T.; Ghanem, B. SoccerNet: A Scalable Dataset for Action Spotting in Soccer Videos. *arXiv* **2018**. arXiv:1804.04527.
8. Li, Y.; Zhang, X. SiamVGG: Visual Tracking using Deeper Siamese Networks. *arXiv* **2019**. arXiv:1902.02804.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).