BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY
Excel@FIT 2025
http://excel.fit.vutbr.cz

Automated Tracking of Players and Ball Movement in Football Video Footage

Maksim Samusevich

Abstract
Today, professional sports represent a huge business, where teams are constantly seeking new ways to optimize their performance. In football, large coaching staffs increasingly rely on modern technologies, including computer vision, to analyze player behavior and improve tactics. This work focuses on the automated analysis of football videos by detecting and tracking players, referees, goalkeepers, and the ball using computer vision techniques. The project builds upon an existing pipeline and extends it by adding improvements and new modules for detailed match analysis. Our approach combines object detection, multi-object tracking, homography transformations, and radar-based visualization.

*xsamus00@stud.fit.vutbr.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

Football has evolved into a global industry where tactical precision and continuous improvement are essential for success. Coaching staffs are looking for ways to better understand team dynamics and player behavior. Automated tracking of players and the ball offers detailed data that helps coaches identify mistakes and refine strategies. In this work, we focus on detecting and tracking key objects (players, referees, goalkeepers, and the ball) in regular broadcast footage and projecting their positions onto a top-down radar view using homography. Our solution maintains player identities, distinguishes between teams, and estimates smooth trajectories. In addition, we introduce pressing intensity analysis and generate heatmaps to visualize player activity across the field.

Currently available commercial systems, such as Catapult [1], KINEXON [2], and Mediacoach [3], offer highly accurate tracking of players and the ball. These systems often rely on GPS sensors or specialized multi-camera setups and are widely adopted by professional clubs in top football leagues. However, they are expensive and require complex infrastructure, making them impractical for smaller teams or grassroots projects.

2. Proposed Solution

Our system processes football match videos in several steps. First, two separate object detection models are used — one specialized for detecting the ball and another for detecting players, referees, and goalkeepers, both based on YOLOv8 [4]. Once the key objects are detected, we apply tracking algorithms (BYTETracker [5] and BoT-SORT [6]) to associate detections across frames. To distinguish players by teams, we extract image crops of each player from the frame and pass them through a SigLIP[7] embedding model, which generates feature vectors capturing visual appearance. These embeddings are then reduced in dimensionality using UMAP [8] to make clustering more stable and efficient. After dimensionality reduction, we apply K-Means [9] clustering to automatically group players into two teams based on their visual similarity, without any manual labels.

Since goalkeepers often wear jerseys of different colors, we use a custom algorithm to assign each detected goalkeeper to the nearest team. The algorithm matches each goalkeeper with the closest players based on position and team distribution, ensuring that goalkeepers are correctly associated with their teammates for accurate tactical analysis.

The next step after detection and tracking is projecting the player and ball positions onto a standardized radar view. For this, our system uses a dedicated model to detect key points on the football pitch, such as corners and penalty areas. By matching these detected landmarks with the known real-world dimensions of the field, we compute an accurate homography transformation. This transformation ensures that all player and ball coordinates are mapped consistently onto a top-down view of the pitch, scaled to real measurements.

3. Real-Time and Post-Match Analysis Modes

Our system is designed to operate in two different modes: real-time and post-match analysis.

In **real-time mode**, object detections and tracking are performed frame-by-frame with minimal delay. This configuration is optimized for live use, providing immediate feedback during a match.

In **post-match mode**, additional processing is applied to improve the completeness and quality of the data. We perform ball trajectory interpolation, filling gaps where the ball detection might temporarily fail, especially during fast or aerial passes. Corrections for aerial ball trajectories are also applied, ensuring that the ball draws the accurate trajectory on 2D radar.

Furthermore, post-match mode enables extended analytics. It supports the estimation of player speeds and movement directions, calculation of pressing intensity for individual players, and the generation of heatmaps that visualize player activity across different zones of the field.

4. Results

The biggest challenge in our system was achieving reliable ball tracking, as the ball is small, fast, and often occluded. We compared three trackers — BYTETracker [5], DeepSORT [10], and BoT-SORT [6] — focusing specifically on ball trajectories. BoT-SORT achieved the best results, reducing the number of frames without ball detection to just 10.8% and maintaining longer, more stable tracks compared to others. To further boost performance, we introduced a dedicated one-class model for ball detection, which improved frame recall from 49% (multi-class model) to 86% and significantly reduced long gaps without detection. In post-match mode, we apply interpolation to fill missing ball positions and correct aerial passes by detecting linear segments using RANSAC [11], finding peaks in ball acceleration, and adjusting the trajectory based on interpolated paths. .

Acknowledgements

I would like to thank my supervisor Ing. Marek Vaško for his support and valuable advice during the writing of this work.

References

[1] Catapult sports performance analysis technology. Accessed: 2025-04-26.

[2] Kinexon - precision tracking for sports teams. Accessed: 2025-04-26.

[3] Laliga mediacoach - advanced sports analytics. Accessed: 2025-04-26.

[4] Glenn Jocher et al. Yolov8: Ultralytics next-generation object detection models. *https://github.com/ultralytics/ultralytics*, 2023.

[5] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Chi Weng, Zehuan Yuan, and Ping Luo. Bytetrack: Multi-object tracking by associating every detection box. In *European Conference on Computer Vision (ECCV)*, 2022.

[6] M. Aharon et al. Bot-sort: Robust associations multi-pedestrian tracking. In *European Conference on Computer Vision (ECCV)*, 2022.

[7] Hugo Touvron et al. Siglip: Scaling and generalization improvements for vision-language models. *arXiv preprint arXiv:2210.08901*, 2022.

[8] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[9] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1. University of California Press, 1967.

[10] Nicolai Wojke, Alex Bewley, and Dietrich Paulus. Deepsort: Simple online and realtime tracking with a deep association metric. arXiv preprint, 2017. arXiv:1703.07402.

[11] Martin A Fischler and Robert C Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.