Football players detection with HOG, CAMShift and Kalman filters

Lorenzo Gandini
University of Trento
Signal, Image and Video
a.y. 2023/2024

1 Introduction

The goal of this project was to create a software able to recognize football players in a choosen video and track them by drawing colored bounding boxes around them. The color of these boxes should matches the team that each player belongs to. To achieve this, I have developed an algorithm that uses three computer vision techniques:

- **Histogram of Oriented Gradients (HOG)**
- **Continuously Adaptive Mean Shift (CAMShift)**
- **Kalman Filter**

The solution proposed was developed in Python, mostly using the **OpenCv** library since it contains all the functions needed to implement the aforementioned methodologies. The choosen video was sourced from a Bundesliga dataset available on Kaggle [4]. It starts with an attacking action, rather than a static arrangement of players, presenting a **more dynamic and challenging** scenario for the project. Furthermore the camera framing is not fixed, since it pans across both axes and varies in zoom and focus, introducing further complexity to the task. Section 2 introduces the core algorithms of the project, explaining how they works and how they can solve the problem presented. Section 3 delves deeper, detailing my project algorithm's specifics and its application. The final section reflects on the algorithm's limitations, discussing critical choices and potential solutions.

2 Related Methodologies

To achieve the goal outlined in the previous chapter, the pipeline of the algorithm was structured in two main phases:

2.1 Player Detection

For this goal I decided to use the **Histogram of Oriented Gradients (HOG)**. This technique, widely recognized for its effectiveness in pedestrian detection (Dalal et. al [3] and related works), serves as the foundation for identifying football players. Following n the successful applications with basketball player detection using the Daimler dataset [2], the HOG method was adapted for accurate football player identification in this project. It functions by evaluating the direction (orientation) and strength (magnitude) of image edges. The features extracted are then matched against models in the library dataset, enabling the algorithm to differentiate human figures from other objects effectively.

2.2 Tracking the Players

After successfully detecting the players, their positions are tracked frame-by-frame using the **Continuously Adaptive Mean Shift (CAMShift)** algorithm. Unlike the traditional Mean Shift algorithm, which relies on a fixed-size search window, CAMShift dynamically adjusts the window's size and orientation to match the players' movements and size changes throughout the video. This adaptability makes it particularly suitable for tracking in sports videos, where players frequently change speed and direction. It operates focusing on the color distribution within the specified search window. A color histogram is calculated for the search window, and back projection is used to create a probability distribution map of the target's appearance in the current frame. CAMShift then seeks the maximum density area within this distribution, adjusting the search window's size and position to ensure continuous and optimal tracking. The process iterates, adapting to changes in the target's movement and appearance.

After successfully detecting the players, their positions are tracked frame-by-frame using the **Continuously Adaptive Mean Shift (CAMShift)** algorithm. This method was selected over the traditional Mean Shift due to its ability to dynamically adjust to the players' movements and changes in size during the video, rather than relying on a fixed bounding box. Each set of coordinates identified by the HOG technique defines a Region of Interest (ROI), which is then fed into CAMShift to ensure continuous tracking of the players across video frames.

The Kalman Filter enhances tracking's robustness and efficiency by estimating the future state of the tracked object. It supports CAMShift [1] by providing a predictive model that forecasts the players' future positions by analyzing their current and previous locations, in order to compensates for the algorithm's momentary inaccuracies, especially in cases of occlusion, overlapping or erratic movements. By integrating the Kalman Filter, the system can maintain a consistent track of the object, using the filter's state prediction to correct or confirm the tracking window determined by CAMShift.

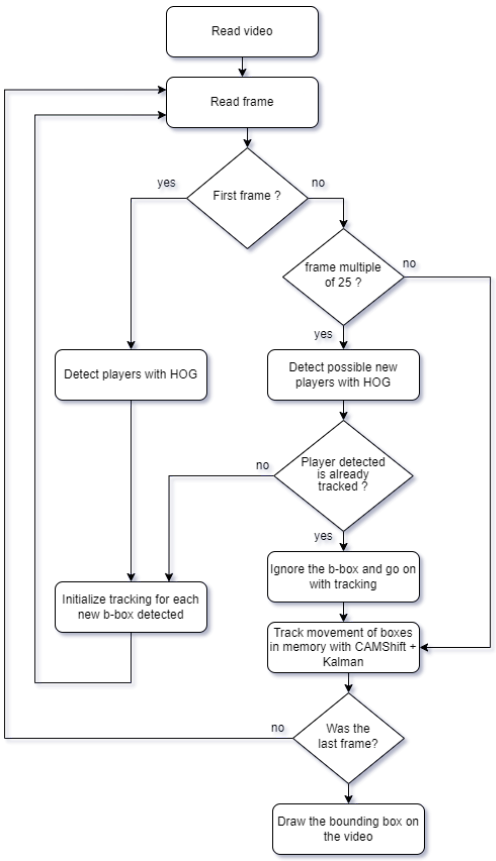


Figure 1: Workflow of the algorithm

3 Solution proposed

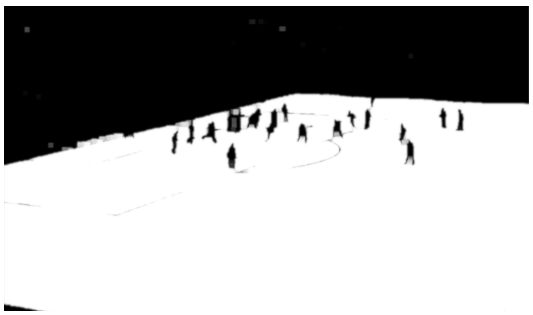
This section delves into a detailed exposition of the proposed solution, explaining the algorithm's workflow of the project. Based on the methodologies outlined in the papers by Cheshire et al. [2] and the combined use of CAMShift and the Kalman Filter [1], my approach adopts a frame-by-frame analysis. This strategy entails executing distinct operations based on the frame number. The operational sequence and decision-making logic of the algorithm are represented in the block diagram illustrated in Fig. 1, providing a visual overview of the system's flow.

3.1 Player detection with HOG

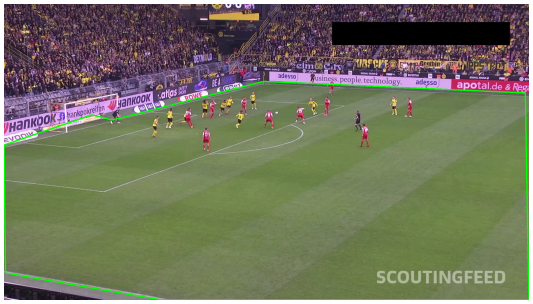
The detection process is initiated on the first frame and subsequently every 1 second (equivalent to 25 frames). This periodic re-detection is strategically designed to account for instances where not all players are initially detected, which may occur due to occlusions by other players or limitations in the detection algorithm. This approach ensures a more comprehensive identification of players throughout the video, significantly enhancing the tracking accuracy by periodically updating the positions.

- **Image Preprocessing**: The method for the identification of players begins by isolating the football pitch in order to speed up the detection process avoiding all the crowd of the stadium. A binary mask, created using the green HSV color space, identifies the green pitch, excluding irrelevant areas. We apply a Gaussian blur to this mask to reduce noise, and also a morphological opening (erosion followed by a dilatation) (Fig. 2a)to remove imperfections and artifacts,obtaining the area highlighted in the Fig. 2b . Dilation of the pitch's contours ensures coverage of peripheral areas, capturing also players close to the sidelines. In the end, this phase returns an image ready to facilitate player identification with HOG.

- **Applying HOG Descriptor**: The HOG descriptor is the central part of this process. Before applying HOG, the ROI obtained with the preprocessing, undergoes adjustments in luminance and saturation, with a contrast enhancement to prepare the image for gradient detection. The HOG descriptor, adapted from pedestrian detection models (Daimler dataset), scans the ad-



(a) Pre-processing operation for court detection



(b) The resoulting court detected with the green line

justed ROI. It divides the image into small connected regions (cells), computes a histogram of gradient directions for the pixels within each cell, and normalizes these histograms across larger, overlapping blocks for improved accuracy. This method effectively isolates the distinct silhouettes of football players against the varied background of the pitch. Each area of interest is identified by the x and y coordinates of the top-left corner of the rectangle, along with the rectangle's width and height. All these regions are stored and returned into the final phase of player identification.

- **Non-Maxima Suppression**: Following the HOG descriptor's identification of potential player locations, overlapping detections are very common. Non-maxima suppression processes these detections, retaining only the most significant bounding box for each player. This critical step filters out redundant boxes, ensuring a clear and singular representation for each detected player. These boxes delineate the positions of all detected players within the specific image. The coordinates will be stored in a list and passed as input to the tracker.

3.2 Keep Track of Players

In every video frame, the bounding boxes stored in memory are updated through the combined use of CAMShift and Kalman filters. As illustrated in Fig. 1, these bounding boxes may comes from HOG detection or from the result of the following tracking processes.

Each detected bounding box is defined by a unique ID, linking it to a specific Kalman filter. New bounding boxes trigger the initialization of new filters, to booster and refine CAMShift's tracking efficacy as discussed in section 2.2

- **Pre-processing of ROI**: The coordinates defining a bounding box determine the Region of Interest (ROI) for individual player analysis. This ROI undergoes Gaussian Blurring to mitigate noise, a conversion from RGB to HSV color space, and the extraction of a color histogram-based back projection, which then serves as input for CAMShift.

- **Prediction of State**: The Kalman filter associated with each bounding box's ID predicts the future position of the x and y coordinates.

- **Apply CAMShift and compare Results**: With the pre-processed ROI, CAMShift attempts to track the object. The reliability of a bounding box's tracking process is assessed by its area; an area below a predefined threshold signals unreliability. This criterion addresses CAMShift's occasional conflation of multiple players or extraneous objects into a single entity (like advertising banners) as shown in Fig. 3 during some test. Conversely, the Kalman Filter contributes by forecasting distinct trajectories for each entity. Unreliable areas prompt an update of coordinates using the Kalman Filter's predictions for that specific frame. In this way we can give a more accurate and consistent player tracking system while minimizing discrepancies.

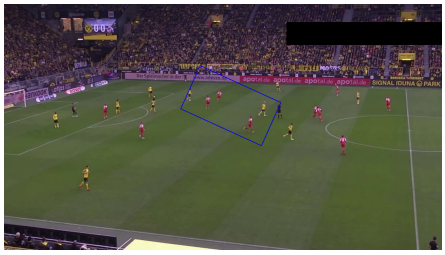
3.3 Incorporating New Players into Tracking

As previously outlined, the player detection process using HOG runs every 25 frames to identify any new players that may have entered the scene. Given that this detection could potentially recognize players already being tracked, it was necessary to develop a matching algorithm to avoid multiple useless tracking.

The matching process involves comparing each newly detected bounding box from HOG against each existing bounding box in the current frame's tracking list. This comparison is based on two key criteria:

(a) Frame 1

(b) Frame 150



(c) Frame 200

(d) Frame 240

Figure 3: Tracking error case with CAMShift due to the proximity of players of the same color

- **The Euclidean distance** between the centroids of the bounding boxes, which must fall below a specified threshold to be considered a match, indicating close proximity.
- **The Intersection over Union (IoU)** of the boxes, which must be over a defined threshold, suggesting a significant overlap and, therefore, a high likelihood that both boxes are tracking the same player.

When both conditions are met, it is concluded that the bounding boxes correspond to the same entity, and the position is updated to reflect the most recent tracking data. Otherwise, if a bounding box detected by HOG does not match any existing tracking box (signifying the presence of a new player) a new tracking object is initiated for this player, mirroring the process used in the first frame.

3.4 Detection of team colors

Upon detecting a box and confirming its tracking reliability, the frame portion within the coordinates undergoes analysis. This process calculates the average, median, and mode of colors inside the ROI. It then measures the Euclidean distance from standard RGB values for yellow and red (since they are the team colors inside the video). The bounding box's color is decided based on which color aligns with at least two of these statistics, ensuring in this way an accurate team representation.

3.5 Filtering False Positive

During the development phases of the project, as will be discussed in the last section of this report, shadows on the field, the video watermark, or other points of image inconsistency were often mistakenly identified as players. Filtering operations were deemed necessary to avoid unnecessary resource usage during tracking phases, in order to avoid a situation as shown in Fig. 4 . Consequently, two filters were implemented:

- **Coordinates** : This check is performed 5 frames after the identification of potential new players, allowing the tracking to operate correctly and subsequently filtering out any bounding boxes whose coordinates have remained unchanged for 4 consecutive frames.
- **Area** : An error occurs when a bounding box updates its position based on predicted movements, but its area remains unchanged because CAMShift fails to detect an object to track. Despite positional changes suggested by the Kalman filter, if CAMShift doesn't adjust the bounding box size due to the lack of a recognizable object, the area stays constant. This discrepancy indicates a tracking error, leading to the removal of such bounding boxes before the final rendering. So all the bounding boxes whose area remains constant for more than 5 frames are removed.

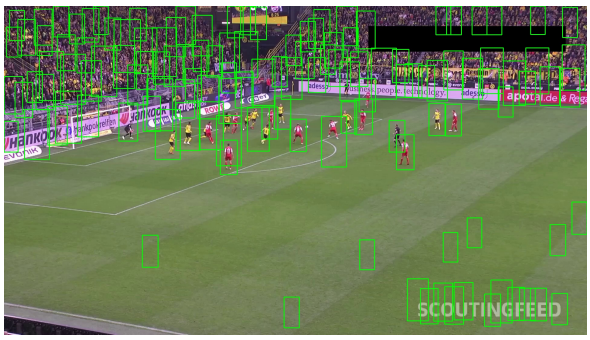


Figure 4: Case of detection without filtering the crowd and false positive from court's shadows

3.6 Drawing of bounding boxes

During all the processes described before, the bounding boxes are added to a list with the following structure:

```
"id": id of the bounding box,
"frame": number of the frame,
"coords": {
"x": x coord,
"y": y coord,
"w": width,
"h": height
},
"color": [RGB Values]
```

At the end of the identification and tracking process, as a last operation, bounding boxes unchanged for at least 5 frames are removed to eliminate inaccuracies. The cleaned list of boxes is then used to create an updated version of the video, with squares around players.

4 Critical Aspects

The provided solution brings some inefficiencies and limitations influenced by various factors:

4.1 Type of Video

As mentioned in the first chapter, the complexity of the video presented several challenges and observation that can change a lot the quality of the result:

- Optimal outcomes are most achievable when the initial frame clearly showcases all players well-distributed across the court, like the tipical positioning seen at the beginning of a match. In such arrangements, players are not overlapped, and have sufficient space that surrounds each individual, simplifying the initial detection process. This clear separation significantly aids in establishing reliable bounding boxes, which in turn, facilitates more accurate and effective tracking.

- The video's camera, intended for streaming rather than analytical purposes, introduced complications. Variations in luminance affected HSV value consistency, complicating the distinction between colors like yellow and green. This often resulted in the misidentification of court shades as players.

- Camera movement across three axes and changes in focus complicated tracker tuning. The prediction of player positions was biased by camera motion, and the apparent size of players varied not only with the prospective, but also with focus adjustments.

Addressing these issues might benefit from implementing adaptive thresholding and environmental condition normalization techniques, such as projection, to improve detection accuracy.

4.2 Chosen Algorithms

The selection of algorithms for player detection and tracking also presents areas for enhancement:

- **Adaptive Thresholding and Environmental Normalization:** Implementing adaptive thresholding can dynamically adjust detection parameters to account for varying lighting conditions and background changes, enhancing detection robustness. Environmental normalization techniques, including geometric projections, could further mitigate the effects of camera movement and focus changes, ensuring more consistent player detection across different video scenes.

- **HOG Model Customization:** While the Histogram of Oriented Gradients (HOG) model provides a solid foundation for player detection, its effectiveness is constrained when based on a generic pedestrian dataset like Daimler's. Customizing this model with a dataset specifically comprised of football players would likely yield significantly better detection accuracy, as it would be more attuned to the unique characteristics of football player silhouettes and movements.

- **Enhanced Kalman Filter Predictions:** Refining the Kalman Filter's predictive model with advanced machine learning techniques could

substantially improve its capacity to forecast player positions, enhancing the overall tracking reliability, especially in complex tracking scenarios.

- **Convolutional Neural Network (CNN) Implementation:** The incorporation of CNNs for player detection and tracking offers a substantial upgrade over traditional methods. CNNs excel at recognizing complex patterns in visual data, making them well-suited for distinguishing players from complex backgrounds, even in the presence of occlusions and varying appearances. Training a CNN with a football-specific dataset could markedly increase the precision of player identification.

- **Advanced Matching Criteria:** Expanding the matching criteria beyond simple Euclidean distance and Intersection over Union, by incorporating shape-based metrics and possibly the insights from CNN analysis, could drastically reduce matching errors. This enhanced approach would allow for more differentiation between players, significantly improving the system's ability to accurately track individual movements.

- **Optical Flow Analysis:** Integrating optical flow analysis can improve tracking accuracy by leveraging the motion vectors of players between frames. This approach is particularly beneficial for predicting player movements and adjusting tracking algorithms in real-time, offering a solution to tracking challenges posed by rapid or unpredictable player actions.

References

[1] Sukhrob Atoev, Akhram Nishanov, and Fakhriddin Abdirazakov. Object tracking method based on kalman filter and camshift algorithm for uav applications, 2021.

[2] Evan Cheshire, Cibele Halasz, and Jose Krause Perin. Player tracking and analysis of basketball plays, 2015.

[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection, 2005.

[4] Saber Ghaderi. Dfl bundesliga - 460 mp4 videos in 30sec + csv, 2021.