Implementation of Football Player Tracking and Performance Analysis Using the OpenCV Library

1st Saúl Leyva Santarén
International Master in Computer Vision
Universidade da Coruña
A Coruña, Spain
saul.leyva.santaren@udc.es

***Abstract*—This study showcases a review and implementation of the method proposed in the paper 'Football Player Tracking and Performance Analysis Using the OpenCV Library'. It evaluates the effectiveness of the solution using the OpenCV Library, aiming to confirm the method's applicability and identify potential enhancements for sports analytics.**

***Index Terms*—Football Player Tracking, OpenCV, CSRT Tracker, Speed Estimation, Sport Analytics**

I. Introduction

Football is currently the most popular sport in the world, and has been attracting new people in the previous years [1]. This growing popularity of the sport is largely due to the improvements in gameplay that have been observed in recent years. Trying to achieve the best team possible, coaches and strategists are always looking for new and creative ways to improve the performance of their players.

However, analyzing a team's gameplay is a very time consuming task and requires a lot of attention to detail. It demands watching numerous hours of matches, not only to extract insights about one's own squad but also to gain strategic advantages by understanding other team's tactics.

Computer vision tools are capable of simplifying the analysis of complex sports data. With the use of sophisticated algorithms and/or machine learning methods, computer vision tasks offer an innovative approach to examine football games with efficiency and accuracy.

Driven by the potential of these technological advancements, this study aims to investigate and assess the methodology described in "Football Player Tracking and Performance Analysis Using the OpenCV Library" [2] by Merzah et al. The suggested solution will be implemented, highlighting its advantages and disadvantages and giving possible improvements to do in the future.

II. Literature Overview

The main objective of 'Football Player Tracking and Performance Analysis Using the OpenCV Library' [2] is to develop a robust system that converts the physical behavior of football players into quantifiable data such as the speed of the player or an activity count. This transformation allows for a more straightforward and analytical approach to understanding the player performance. The development of this solution was made using the tools offered by OpenCV library.

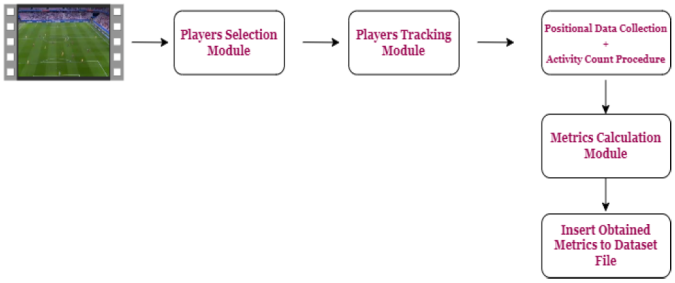

Fig. 1. Model block diagram proposed in the paper 'Football Player Tracking and Performance Analysis Using the OpenCV Library' [2].

The proposed framework is shown in Figure 1. They use a tactical camera video of a football match as the input. In the first frame of this video, the trackers are initialized with the manual selection of the players' bounding box. They decided to select a total of ten players of one of the teams each time, excluding the goalkeeper.

Despite OpenCV offers several trackers to do this task, they decided to use the CSRT tracker [3]. As shown in Figure 2, this tracker works in two main steps: the location and the update step. To achieve excellent performance, it creates a multi-channel feature representation of the tracked item by combining features like HoG with other feature descriptors like color histograms. As a result, the tracker is able to perform a robust tracking of the object. With this method, they track each of the ten players of each team (one team at a time) and save their bounding boxes across the video sequence in a csv file.

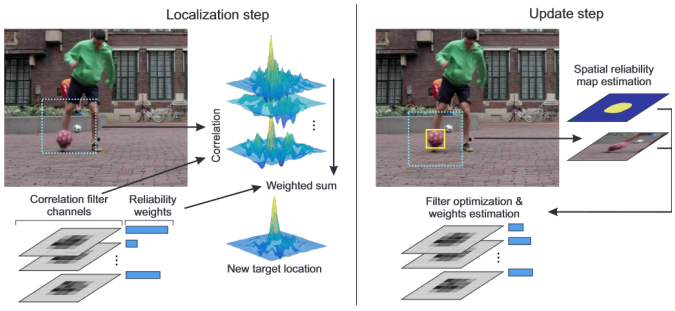

Fig. 2. The CSRT tracking iteration, showcasing the localization step (on the left) and the update step (on the right) [3].

After tracking and storing the players' positional data in the CSV file, the next step involves processing this data to compute several metrics. The initial metric is the Cross-Distance (CD), which represents the sum of the Euclidean distances between the player's positions in two successive video frames (Equations 1 and 2). This calculation provides a direct measure of the total distance covered by a player during the match, offering insights into their physical activity and mobility on the field.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{1}$$

$$CD = \sum_{i=1}^{n} d_i \tag{2}$$

Since the CD is measured in pixels, they transformed this distance in meters. To do so, they obtained the Pixels Per Meter ratio (ppm) (Eq. 3) using the standard drawing scale relationship and the standard size of a soccer field. Once the ppm is obtained, they used the Equation 4 to transform the previously obtained Cross-Distance into meters.

$$ppm = \frac{\text{Distance in image}}{\text{Real distance}} \tag{3}$$

$$dm = \frac{\text{Cross-Distance (CD)}}{ppm} \tag{4}$$

With the real distances calculated, we can obtain the speed of the player given the distance traveled in a certain amount of time, using the Equation 5.

$$\text{speed} = \frac{\text{Distance}}{\text{Time}} = \frac{\text{Cross-Distance}}{\text{Video time}} \tag{5}$$

After that, they obtain the activity count metric. It measures the activity of a player with actions such as passes, touches of the ball, tackles, and player-to-player contact. To do so, they compare the histograms of a region of interest given by the Equation 6 in consecutive frames (Fig. 3). In case there is a deviation in the correlation between them, the activity counter increments in 1.

$$ROI = \text{frame}[y + \text{int}(h * 0.75) : y + h, x : x + w] \tag{6}$$

Finally, to validate the efficacy and robustness of the methodology, they made tests using three distinct videos recorded by a tactical camera, capturing football matches under varying conditions. They tracked one, three, five, seven, and ten players across these videos to evaluate the scalability and accuracy of the tracking system under different levels of complexity.

The accuracy of the tracking results was quantitatively measured using the following equation for accuracy (Eq. 7), which enabled to assess the performance of the tracking methodology across different scenarios.

$$\text{Accuracy} = \frac{\text{Number of correctly tracked players}}{\text{Total number of selected players}} \tag{7}$$

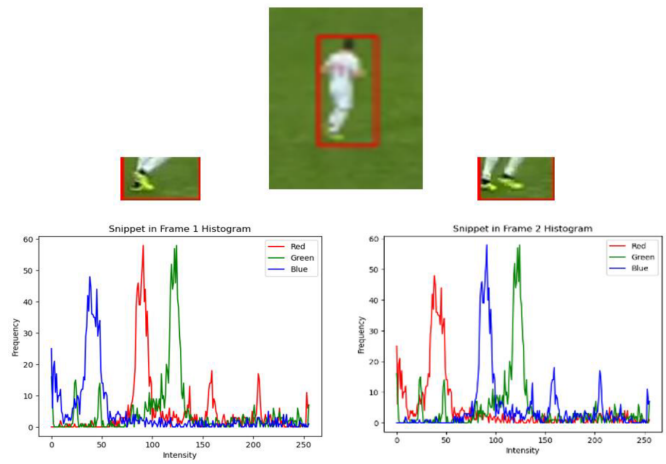

Fig. 3. Histogram comparison in activity count metric [2].

III. Experiment

In this section, an overview of the video dataset utilized and the methodology implemented for our experiment will be delineated.

A. Dataset

The original paper did not provide details about the dataset used, nor the specific football matches from which the videos were sourced. Despite the efforts to locate datasets with ground truth player data and relevant statistics, such as speed or real location in the field, captured with tactical cameras like to those described in the paper, none were found.

It was also very hard to find football matches with the tactical camera, because this view is not the one generally televised. As a result, I had to obtain two football matches from a repository on Mega [4] that contained various matches from the 2022 FIFA World Cup. Specifically, I selected to work with the Spain-Japan and Brazil-Croatia matches because all four teams featured kits with distinct colors.

Given that the matches were nearly two hours in duration, I extracted a single sample from each match of about 20 seconds long approximately, with unique characteristics. Both of the videos have a frame-rate of 30 frames per second. As illustrated in Figure 4, the sequence from the Spain-Japan match predominantly showcased players positioned farther from the camera, resulting in more pronounced perspective distortion, while the video sample from the Brazil-Croatia match featured players closer to the camera.

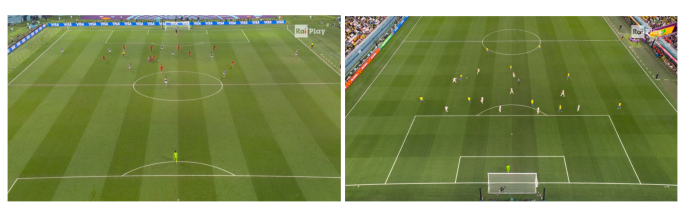

Fig. 4. Frames taken from the Spain-Japan match (left) and Brazil-Croatia match (right) [4].

B. Methodology

In order to perform the tracking of the players, I followed the exact same procedure described in the paper. I first initialized the CSRT trackers in the first frame of the video, and manually select the bounding boxes of ten players of a team, excluding the goalkeeper. I decided to work with 10 players directly, since the results of the paper with less players achieved a perfect accuracy most of the time. Also, I assigned each of the players one different color, to ease the visualization process as depicted in Figure 5. The location of the bounding boxes of each frame was stored in a csv file as in the original paper.

Fig. 5. Frame obtained from the tracking process of Brazil players.

Next, I perform the metric calculation process based on the data obtained from the previous task. The main challenge of this stage was the absence of a clear explanation regarding the methodology for establishing the relationship between pixels and meters, a crucial step in the analysis process. The referenced paper [5] merely mentions providing 'ppm' (pixels per meter) and 'fps' (frames per second) values to the system before conducting speed detection on the input video, without elaborating on the exact procedure. Consequently, I designed my own approach to estimate this conversion of measurement units.

I decided to obtain four known distances within the football field, comprising two horizontal and two vertical lines relative to the camera. My aim is to derive the Pixel-Per-Meter (ppm) ratio for both the horizontal and vertical components to interpolate their values for the correspondent movements of the players in the respective axis.

In the initial frame of the video, I ask the user to select four pairs of points along with their corresponding real distances: one pair for a horizontal line in the background, another for a horizontal line in the foreground, one for a vertical line in the background, and one for a vertical line in the foreground. Considering that the camera in the obtained videos exhibits movement and zooming, I employed the Lucas-Kanade optical flow method [6] to track the position of these points on the field.

Using Equation 1 to calculate the pixel distance between each pair of points, I then applied Equation 3 using the associated real distance to derive four ppm values. As depicted in Figure 6, it is apparent that the further the points are from the camera, the lower the ppm ratio, as fewer pixels are required to span a meter's distance. Additionally, the disparity between the ppm ratios of vertical and horizontal lines is evident, with horizontal lines exhibiting higher values due to the perspective distortion effect. I decided to assign each of this ppm values to their corresponding y axis value of their center position, because the further apart they are (higher position in the image), the more distortion it gets.

Fig. 6. Example of reference points with real life values and the corresponding ppm value for that frame.

For simplicity, each player, based on their position along the y-axis, is assigned two interpolated ppm values derived from Equation 8. In this equation, x represents the y-axis position to interpolate, x_0 and x_1 correspond to the y-values of the two central points of the given vertical/horizontal reference lines, y_0 and y_1 are the real-world assigned to these points, and y is the interpolated ppm value at position x. Accordingly, for any given position, this equation is applied using the two horizontal lines to determine the horizontal ppm, and the two vertical lines for the vertical ppm, for the player located at that y-axis position.

$$y = y_0 + \frac{(x - x_0) \cdot (y_1 - y_0)}{x_1 - x_0} \tag{8}$$

Equation 4 allows us to convert the pixel distance between consecutive frames into real distance along the horizontal and vertical axes by dividing it by the corresponding ppm value (either horizontal or vertical). By doing this, we obtain the actual horizontal and vertical distances traveled. Subsequently, we apply the Euclidean distance formula (Eq. 1) to these distances to calculate the total real distance traveled by the player. This calculation is updated for each frame.

To calculate the speed, we employ a similar approach. I have chosen to sample speed on a second-by-second basis. This involves taking the player's position at a given frame x as the initial position and the position at frame $x + 30$ as the final position. The distance traveled during this interval is computed, and given that 30 frames equate to one second, the denominator in Equation 5 becomes 1. Consequently, in

this context, the distance traveled directly corresponds to the speed.

Finally, the measurement of activity level is equal to the methodology outlined in the original paper (Fig. 3), employing Equation 6 with a minor modification. To accommodate the drawing of bounding boxes, I adjusted the approach by excluding 2 frames from both the sides and the top. This adjustment ensures that the analysis focuses on the relevant area within the bounding boxes, taking into account the differences in the histograms of the region of interest (ROI) between consecutive frames.

IV. Results

In Table I, we can see a comparison between the performance of the proposed method and the experiments of the paper with 10 players. In the paper, three different experiments were conducted. None of them had occlusions in the video. The first experiment achieved a perfect score of 100% accuracy. The second experiment achieved 80% accuracy because it experienced an ID switch and lost track of one of the ten players. Finally, the third experiment had only one lost track, achieving an accuracy of 90%. So, on average, they achieved 90% accuracy in their tests.

On the other hand, the videos chosen in the dataset of this study contained occlusions between players of the same team and players of different teams. In the Brazil-Croatia match, neither of the teams suffered an ID switch or lost tracking, achieving both 100% accuracy. Despite this good performance, the video of Spain-Japan showed lower performance. In the Japan team tracking, there was an ID switch between players of the same team, and in the Spain tracking, there was a player who lost tracking. For these reasons, each of the methods achieved 90% accuracy.

In Figure 7, we can assess the overall quality of the tracking method. We observe in the top row, where the Brazil-Croatia match is displayed, that both teams have their players correctly identified. Conversely, in the bottom row, particularly in the image of the Japan team on the left, we note how the player adjacent to the one with the white bounding box (on the left) lacks a bounding box. This omission is attributed to the player's bounding box being mistakenly associated with his teammate, a consequence of their similar shape and colors, resulting in one bounding box being allocated incorrectly.

Furthermore, in the image of the Spain team, a player on the right side of the image does not have a bounding box. This absence is due to the player moving outside the camera's field of view for a few seconds, rendering the tracker unable to maintain tracking. This scenario highlights one of the primary challenges of this solution: the necessity for the tracked object to remain visible at all times.

The tracking of the position of reference points encounters similar challenges. It's essential to select points that remain visible at all times, as the method cannot maintain tracking if visibility is lost. Additionally, choosing points that are not overlapped by players is crucial. During the testing phase of the keypoint functionality, it was found that the method tends to mistakenly start tracking players who overlap with these points quite often. This susceptibility underscores the importance of careful selection and placement of reference points to ensure accurate and uninterrupted tracking performance.

TABLE I
EXPERIMENT RESULTS COMPARISON

Experiment	Occlusion	Switch ID	Lost tracking	Accuracy
1 (paper)	×	×	×	100%
2 (paper)	×	✓1	✓1	80%
3 (paper)	×	×	✓	90%
Brazil	✓	×	×	100%
Croatia	✓	×	×	100%
Japan	✓	✓1	×	90%
Spain	✓	×	✓1	90%

Regarding the quality of the metrics, both the speed and distance tracking performed well, providing good estimations of the actual speed and distance traveled by the players. It's crucial to mention that these two metrics should be evaluated using static cameras that do not move. This consideration is paramount because a camera's zoom can create artificial speed and distance traveled for the players, as the system detects a change in the players' coordinates. To address this issue in future work, one approach could involve analyzing the movement of the keypoints (which should remain stationary) and applying an equal and opposite vector to counteract the movement caused by the camera. This method could help mitigate the effects of camera motion, ensuring more accurate measurement of player speed and distance.

Finally, the activity measure did not perform very well, as almost all the players exhibited a level of activity similar to the total number of frames. One possible explanation for this could be that the analyzed region of interest (ROI) might have had a very low resolution, meaning that minor changes could significantly impact the overall result. I even tried testing different threshold parameters to be less restrictive with the correlation, but overall, this metric was not very informative.

V. Conclusion

In conclusion, this implementation demonstrates the significant potential of computer vision to streamline the evaluation of player movements and actions, tasks that traditionally require extensive manual effort. While the study achieves notable accuracy in player tracking, it acknowledges certain limitations, particularly in measuring activity and in scenarios involving occlusion or rapid movement.

Tracking methods face challenges inherent in the context of football matches, such as multiple occlusions between teammates. Similarly, tracking key features encounters issues, as at least one player might overlap with key points on the football field at some point during the game.

The metric calculation offers good estimations of speed and distance traveled but struggles with camera movement. This issue, and the problems with keypoint tracking could potentially be resolved by employing stationary cameras that neither move nor zoom during the game.

Furthermore, I recommend exploring alternative methods for measuring the activity levels of players, as the current method did not perform as expected.

Overall, the research lays a strong foundation for further innovation in applying computer vision technologies in sports, opening new paths for enhancing team strategy and player development.

REFERENCES

[1] FIFA, "The Football Landscape," *FIFA Vision Report 2021*, [Online]. Available: https://publications.fifa.com/en/vision-report-2021/the-football-landscape/. [Accessed: 28, March, 2024].

[2] B. M. Merzah, M. S. Croock, and A. N. Rashid, "Football Player Tracking and Performance Analysis Using the OpenCV Library," in *Mathematical Modelling of Engineering Problems*, vol. 11, no. 1, 2024.

[3] A. Lukezic, T. Vojir, L. Čehovin Zajc, J. Matas, and M. Kristan, "Discriminative Correlation Filter with Channel and Spatial Reliability," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6309-6318.

[4] MEGA, "WC 2022 tactical camera," [Online]. Available: https://mega.nz/folder/qUQhibYJ#iCqNtXwVTHlE57Ew8VssMg. [Accessed: 29, March, 2024].

[5] S. Kamoji, D. Koshti, A. Dmonte, S. J. George, and C. S. Pereira, "Image Processing Based Vehicle Identification and Speed Measurement," in *2020 International Conference on Inventive Computation Technologies (ICICT)*, IEEE, Feb. 2020, pp. 523-527.

[6] OpenCV, "Lucas-Kanade Optical Flow," *OpenCV Documentation*, [Online]. Available: https://docs.opencv.org/4.x/db/d7f/tutorial_js_lucas_kanade.html. [Accessed: 29, March, 2024].

Fig. 7. Final results from the implemented method: The top row displays the Brazil-Croatia match, with the Brazil team tracked on the left and Croatia on the right. The bottom row depicts the Spain-Japan match, with the Japan team tracked on the left and Spain on the right.