Kickflow Tracker: A Computer Vision and Deep Learning Approach for Ball Detection, Extraction, and Speed in Football

By: Allen Lau and Rahul Chandani

I. Introduction

Whether it is to gain a competitive advantage or provide a more comprehensive experience to fans, computer vision is being explored for a variety of applications in the sports industry. For example, using computer vision techniques can strengthen our ability to interpret, analyze, and extract useful information from visual data and can provide the opportunity for retrieving instant analysis for decision making. Moreover, these technologies are used to create overlays and visualizations during broadcasts to provide a more immersive experience. These are just some examples in this fast-growing subject.

For this report, the use of computer vision techniques will be explored and applied to the sport of football, also known as soccer. Our focus is on the detection and tracking of a football within videos and the calculation of the real-time speed of it. This implementation, called Kickflow Tracker, can then be applied to applications like film analysis and broadcast overlays, among other potential use cases.

II. Problem & Objective

There is a wealth of information available in videos and images that professional football clubs at all levels could utilize to gain key insights in areas like players statistics, tactics analysis, and injury prevention. Organizations who choose to not use computer vision will not be able to reap benefits like enhanced perception and insight on data and real-time analysis. In fact, without computer vision technologies, these groups will be limited in the statistics they can track efficiently, quickly, and accurately.

Kickflow Tracker aims to provide organizations with an easy-to-use application for ball tracking and speed estimation. The proposed pipeline for this goal of ball detection, tracking, and instantaneous speed calculation is the following: (i) video processing is used to extract frames from the video input and apply grayscale and blur filters; (ii) the Hough Transform and Channel and Spatial Reliability of Discriminative Correlation Filter is used for ball detection and tracking, (iii) optical flow with a calibrated camera is used to estimate the instantaneous speed of the ball; (iv) a GUI will display the results to the user.

III. Related Work

There have been many studies for player and ball detection and tracking which primarily utilize deep learning techniques. For instance, common deep learning models for this application include You Only Look Once (YOLO), Single Shot MultiBox Detector (SSD), and Variational Autoencoder (VAE). One implementation that performed particularly well consisted of a two part approach: (i) YOLOv3 to detect and classify objects like the player, ball, background; (ii) SORT algorithm, using Kalman filtering, to track the object [1]. An example of a non-deep learning approach utilized a setup of multiple fixed cameras to capture the playing field. Image differencing is used to detect the objects and the Kalman filter is used for tracking. This filter results in various object paths being detected, and a multitude of features like the ground plane velocity, longevity, normalized size, and color is used to assign a likelihood to find the path that

is most likely to be the ball [2]. Among these various techniques of tracking in football, certain difficulties are common. For instance, complex occlusions, similar appearance of objects, unpredictable movement, unstable camera motion, and motion blur can contribute to difficulties in the accurate detection and tracking of objects on the football field [3]. With Kickflow Tracker, some of the considerations and methodologies mentioned above will be used in conjunction with speed estimation from optical flow to extract additional information beyond object detection and tracking.

IV. Methodology

A. Ground Truth

To provide a basis of comparison, a controlled environment with a fixed camera capturing the scene was established. The device used with the main camera of an iPhone 14 Pro. Measurements of the scene, as seen in Appendix A, were taken such that the ground truth average speed could be calculated with the following formulas, shown in the below Figure 1:

Figure 1: Speed Calculation for Ground Truth

$$\Delta Time = \frac{EndFrame\ \# - StartFrame\ \#}{Frames\ Per\ Second}$$

$$Speed = \frac{\Delta Distance}{\Delta Time}$$

More specifically, for each trial, where the ball was kicked across the scene in view of the fixed camera, the distance the ball traveled was exactly 1.8288 meters (m). The average speed of the ball can be computed from the number of frames it took to move from point A to point B (over a distance of 1.8288 m) and frame rate of the camera used, or 60 frames per second (fps). The below Table 1 depicts the trials taken and the ground truth speed calculated, which will be used for evaluating the performance of the speed estimation from Kickflow Tracker:

Table 1: Ground Truth Speed

Ground Truth Speed (distance traveled = 1.8288m)		
Trial Number	Number of Frames from Point A to B	Calculated Average Speed (m/s)
1	24	4.57
2	23	4.77
3	20	5.48
4	25	4.39

B. Kickflow Tracker Method

The first step performed is camera calibration. The calibration pattern is a 5x7 checkerboard pattern printed on the U.S. letter (8.5in x 11in) paper. 46 images of the calibration pattern in various positions and orientations in the scene were taken with the fixed camera, as seen in Appendix B. Next, the world and image points of the pattern were found and defined and camera calibration, using the Direct Linear Transform (DLT) algorithm was completed. This results in the following intrinsic parameter matrix:

$$M_{int} = \begin{bmatrix} 3796.232 & 0.000 & 544.075 \\ 0.000 & 2877.279 & 959.079 \\ 0.000 & 0.000 & 1.000 \end{bmatrix}$$

It is important to note that for the purposes of this experiment, distortion parameters are assumed to be zero. The manufacturer listed camera parameters like the focal length (24 mm) are also noted. These will allow us to relate the 2D image coordinate system to the 3D world coordinate system; therefore, allowing us to take the optical flow outputs of pixels/time to meters/second.

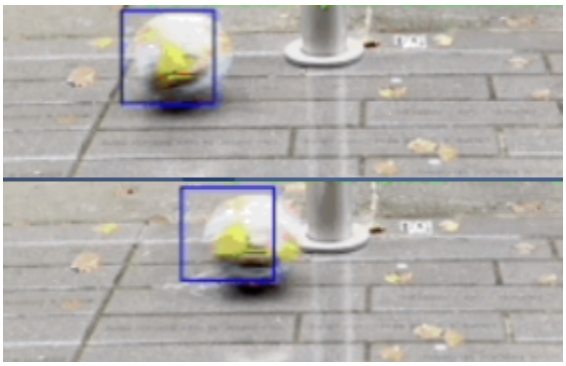
For ball detection, each image frame is preprocessed by applying a grayscale operation and then applying a 9x9 gaussian kernel to blur the image. The object detection approach taken is the Hough Transform (CHT), which can be used to detect circles in the image frames. On a high level, the Hough transform works by creating a parameter space to represent all the possible circles in terms of the centers and radii. Next, a voting strategy and threshold are used to determine the best matches for what is likely to be a circle in the image. Once the ball has been detected, the local window for which the ball is located in the image is used as the input for the subsequent step of object tracking. The below Figure 2 illustrates the ball detection on an example image frame:

Figure 2: Ball Detection via CHT

For object tracking, the Channel and Spatial Reliability of Discriminative Correlation Filter (CSRT) algorithm is used. This is a non-deep learning approach that utilizes a correlation

filter based on the appearance of the target object in the initial frame. This initial frame input is the local window obtained from the prior ball detection step. The tracking of the object is achieved by using the correlation metric to estimate the object's position in subsequent frames of the video. The below Figure 3 is an illustration of the CSRT algorithm tracking a ball across sequential frames:

Figure 3: CSRT Ball Tracking btw. Sequential Frames

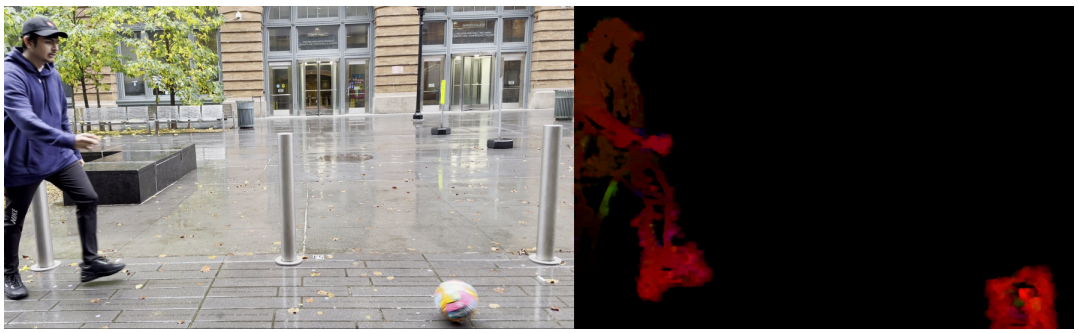


After successfully detecting and tracking the ball in the image frames, optical flow is used to calculate the speed of the ball. The optical flow equation, as seen below, measures the apparent motion of pixels between consecutive frames:

$$\frac{dI}{dx}u + \frac{dI}{dy}v + \frac{dI}{dt} = 0$$

Solving for u and v gives the displacement of the pixels over time (dx/dt, dy/dt). For this application, the Farneback Optical Flow algorithm is used, where we can estimate the dense motion vectors between consecutive frames where the signed-pixel displacement over time can be depicted in an RGB image. The intensity of the pixels represents the magnitude of the displacement over time and the color represents the direction of motion. The below Figure 4 shows an example frame where the ball and man is moving to the right:

Figure 4: Visualization of Optical Flow in RGB Image (Right)

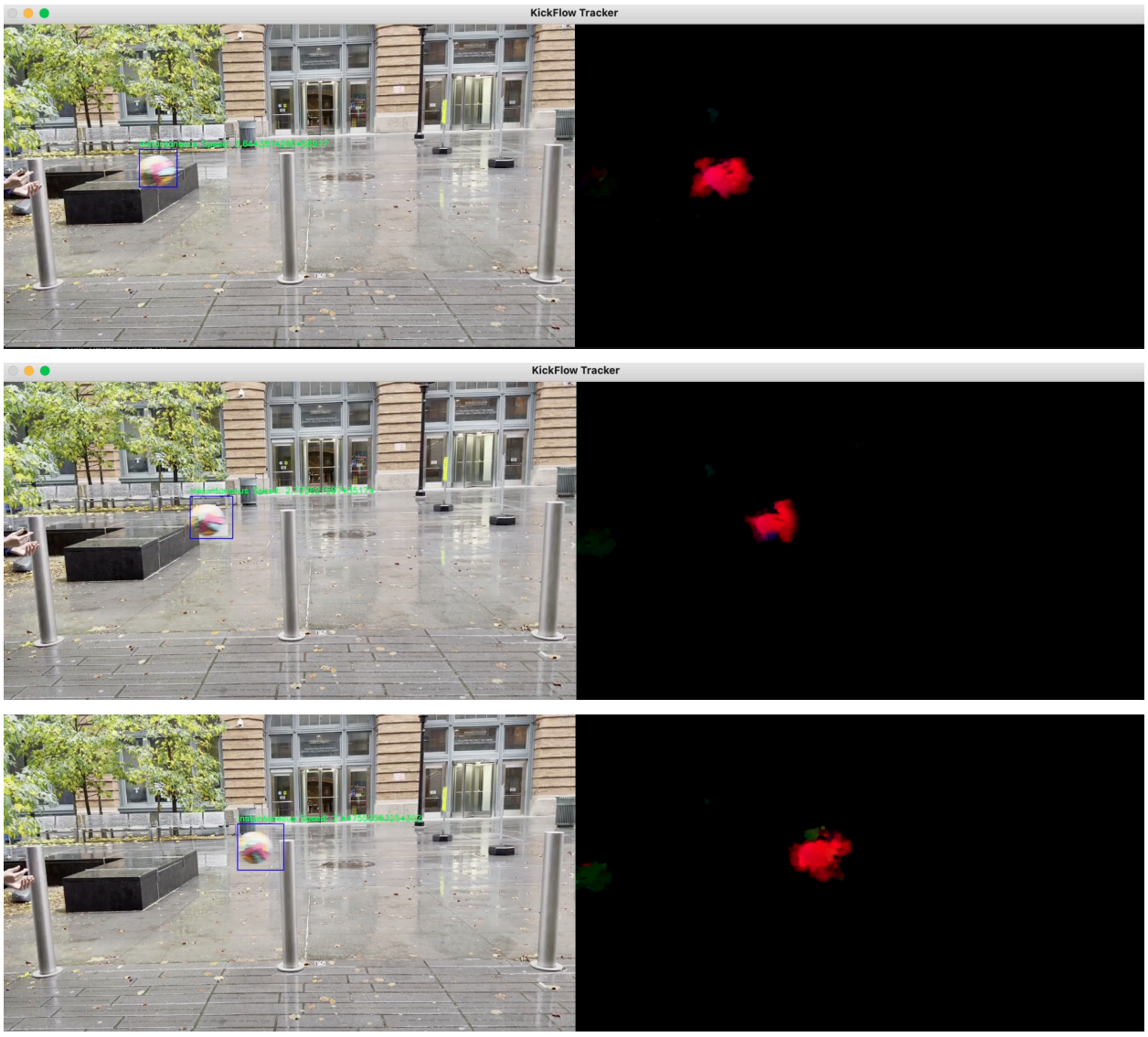


It is then possible to convert from pixels over time to meters per second using the intrinsic parameters of the camera system.

V. Results & Analysis

We package the above technologies into a GUI, where the user can initiate the ball detection and tracking. The ball is automatically bounded by a rectangle with an accompanying instantaneous speed that is displayed. The below Figure 5 are screenshots of the GUI over many different frame examples:

Figure 5: Kickflow Tracker, Displaying Detected/Tracked Ball and Optical Flow Derived Speed



Since we have the instantaneous speed for the entire path of the ball, we will calculate the average speed by only extracting the instantaneous speed between the defined markers used for

the ground truth data points and computing the average. As a result, it is possible to compare the estimated average speed from optical flow to the ground truth average speed, computed using the equations in Figure 1. The below Table 2 depicts the evaluation of the speed calculation:

Ground Truth vs. Estimated Speed via Optical Flow			
Trial Number	Ground Truth Average Speed (m/s)	Optical Flow Derived Average Speed (m/s)	Error (m/s)
1	4.57	3.86	0.71
2	4.77	3.36	1.41
3	5.48	4.02	1.46
4	4.39	4.29	0.10

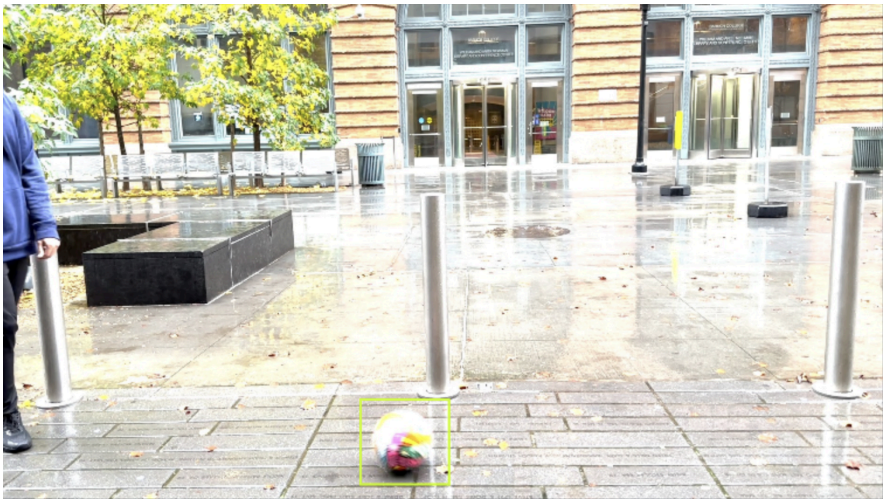
The RMSE Error is computed to be 1.08 m/s with a standard deviation of 0.65 m/s. Although the estimation is not exact, we can conclude that the Kickflow Tracker speed estimation does reasonably well.

VI. Discussion

Although the current implementation of Kickflow Tracker was able to successfully detect, track, and estimate the speed of the ball, the Hough Transform method of ball detection is understood to be unstable. This is especially true with the complex scene of a football game due to the aforementioned issues like complex occlusions and high motion blur. As a result, the primary improvement area for this application is replacing the Hough Transform with a more robust algorithm. This algorithm is chosen to be YOLO, due to its high performance in more complex scenarios and problems.

To train the YOLO model, manual annotation to select the ball in certain frames is completed. These annotated frames serve as the input for training the YOLO model. It is important to note that for this experiment, only annotations of the ball will be used and the model will not be tracking players. Upon successful annotation, the YOLO model takes all the annotated images and generates weights of the model. The below Figure 6 shows the model successfully detecting the ball in our controlled environment, with a higher degree of accuracy and precision than the Hough Transform and CSRT method:

Figure 6 : YOLO Detection of Ball in Controlled Environment



The next logical step is to apply the Kickflow Tracker pipeline to a more realistic football scene, where the fixed camera would be located high and far away from the field of play and would look onto the playing field. The YOLO model was able to track certain sequences of frames successfully, where the Hough Transform and CSRT were not able to, like in the below Figure 7. Unfortunately, the preliminary results indicate that further improvements in training via annotating more training frames to result in a better performing YOLO model is required. Figure 8 illustrates an example where the YOLO model was not able to successfully detect and track the ball. This improvement in the training of this deep learning model will be a primary next step for the team's future work.

Figure 7 : Successful Detection of the Football in Live Match

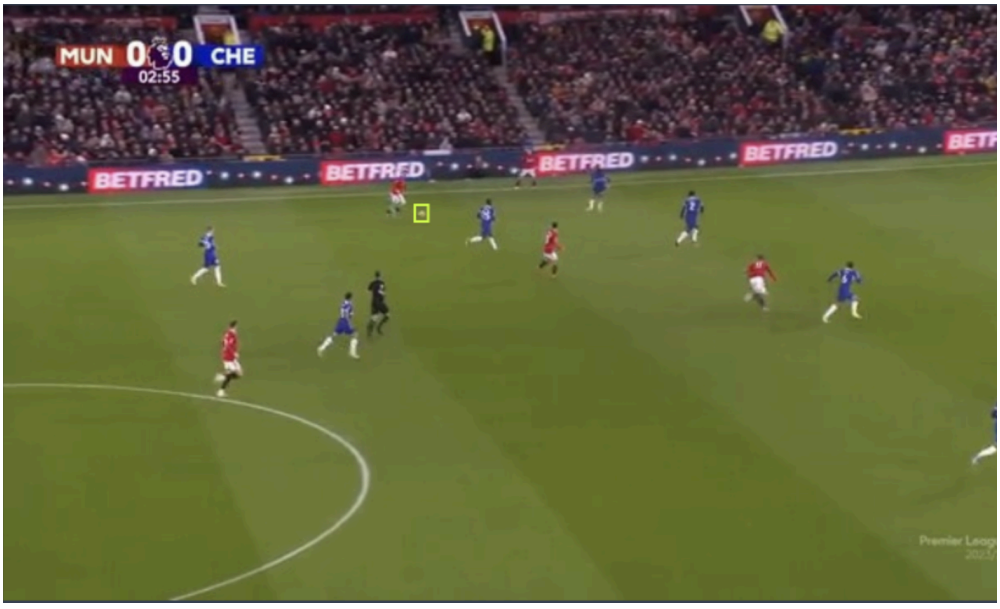


Figure 8: Unsuccessful Detection of the Football in Live Match

VII. Conclusion

Kickflow Tracker aims to be an easy-to-use tool for Football organizations, ranging from high schools to professional teams. It provides the ability for automatic ball tracking while providing quantitative information about the real-time speed of the motion of the ball. However, we recognize that there are improvements that can be made to this system. For instance, improved camera calibration will help decrease the errors of the speed calculation. Additionally, more robust logic for object detection and tracking that utilize state-of-the-art deep learning approaches will help in accurately finding and tracking the ball when there is excessive occlusions, motion blur, or other complexities in the visual information. Lastly, there are many more opportunities for extracting more information in these videos and images that should be explored. For example, the velocity (speed and orientation) of the motions and the biomechanics of the players kicking can provide vast amounts of insight for players, coaches, and staff.

VIII. References

[1] Banoth, N, Farukh H. (2022, March 22). YOLOv3-SORT: detection and tracking player/ball in soccer sport. Retrieved from https://doi.org/10.1117/1.JEI.32.1.011003.

[2] Ren J, Orwell J, Jones GA, Xu M. (2009 May). Tracking the soccer ball using multiple fixed cameras. Retrieved from https://doi.org/10.1016/j.cviu.2008.01.007.

[3] Naik, BT, Hashmi, MF, Bokde, ND. (2022, April 22) A Comprehensive Review of Computer Vision in Sports: Open Issues, Future Trends and Research Directions. Retrieved from https://doi.org/10.3390/app12094429

IX. Appendix

A. Measurements of the Controlled Environment

B. Camera Calibration Pattern Taken in Various Positions in the Scene