UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering, Mathematics and Computer Science

Master thesis

Interaction Technology

FIFA 2022 GAME ANALYSIS USING MACHINE LEARNING & COMPUTER VISION

Shalini Singh

Supervisors

Dr. Mannes Poel

Dr. Guido Bruinsma

Acknowledgment

This master’s thesis is the primary last step towards achieving my Master of Science (M.Sc.) degree in Interaction Technology. Working on this thesis has helped me to extend my knowledge, and my skills in machine learning and computer vision and to further understand the challenges of implementing these technologies in a real application.

I am deeply indebted to my supervisors, Dr. Mannes Poel and Dr. Guido Bruinsma, for being great mentors, for guiding me in every weekly online meeting, and for always taking the time to give me valuable feedback regarding the work that I have done.

Finally, this journey would not have been possible if it were not for the unconditional love and support from my close ones; my husband Abhishek and my son Viraj.

Enschede, May 2022
Shalini Singh

ABSTRACT

The eSports games are gaining momentum among the general public and there is a need to analyze the eSport games using machine learning and computer vision techniques. This helps in understanding the tactics and strategies of the players while providing necessary input useful for improving the quality of the play. The same methods can also support the analysis of the real football games in which API is not available due to a lack of resources.

The objective of the present thesis is to analyze FIFA 2022 eSports football games for gaining knowledge related to the strategy and tactics of the eSport player. For this purpose, accurate detection and continuous tracking of the ball and players is important. Furthermore, the ball possession has to be continuously identified during the analysis. Due to the unavailability of API for eSports football games, the desired information related to the tracking of the ball and players is not available. Therefore, the computer vision with the machine learning approach is used for the detection and tracking of the football which makes it challenging as it is a very fast-moving, small object. Due to this motion, its shape and size can vary in the recorded videos. Furthermore, occlusion because of players makes it difficult to track continuously. Similarly, the tracking of players has a challenge in terms of occlusion by other players. For accurate estimation of players and ball location, the playfield has to be mapped which has challenges because of the camera set up which includes single stationery and a rotating camera. Furthermore, how this extracted data can be utilized for the analysis is also an important component in achieving the objective of the current research.

During this research, three generic models 1) color-based detection and tracking, 2) template-based detection and tracking, and 3) deep learning (Yolov3 model) are selected from the literature study. These models are further developed for tracking the ball and the players from the eSports football videos. It is found that the color-based detection method is the most suitable method for tracking the football in the minimap while YOLOv3 provided the best tracking of players from the bigger field. In the scenarios in which both ball and players are supposed to be detected and tracked, a combined model is utilized. The template-based method has not provided the desired accuracy in detection therefore it was not used in the game analysis.

For the analysis, the field is divided into different zones such as action, mid-field, and attack zones, and the statistics related to ball location in different zones are calculated. In the analyzed video, it is observed that the ball was mostly in the green team's half (~ 60%) which means that it was mostly in the defense mode which is proven by a higher number of goal attempts from the red team in comparison to the green team (4 vs 1). The preferred attacking area was the top part of the attack zone for the red team and the middle attack zone for the green team. Furthermore, the ball path for each goal attempt is also traced from the video. This provided an interesting insight into the strategy of the eSports player from a strategic perspective.

The model was not capable of detecting all the players present on the field, but it was able to detect the players who are close to the ball. Since the eSports player only controls one player at any given moment therefore detection and tracking of players close to the ball is sufficient for understanding the eSport players' game strategy. Even though only one game was analyzed, the developed model was able to provide an interesting insight into the eSports player’s strategy. For a more thorough understanding of the strategy and tactics of the player, several games should be analyzed which can provide patterns of the eSports player suitable for the prediction of their game.

TABLE OF CONTENT

List of Figures

1. Introduction

1.1 Motivation

The eSport FIFA 2022 based football matches are gaining popularity among the general public. Recently FIFA 22 has decided to bring more real-life circumstances to the forefront of the e-game. Similar to the actual sports matches, eSports are also analyzed by the experts for understanding the tactics and strategies of the players and the teams. This understanding can be highly useful in predicting the future performance of the team which can be utilized for improving the defensive/offensive planning of the opposite team. Furthermore, this information is also well sought after by the betting companies or individuals who would like to place their bets based on knowledge rather than an arbitrary guess regarding the match outcome.

The primary difference between an actual and eSport game analysis is that in an actual game we have to investigate each player separately but in an eSport match the whole team is played by only one operator and ultimately we aim to analyze tactics and strategies of only one person which is the operator of the team. Therefore, the analysis focus shifts from a single player to team performance.

1.2 Objective and Problem Statement

The objective of this thesis is to analyze an eSports football match to determine the playing tactics and strategies and tactics of the eSports player. For real football games, APIs are present to collect the data, but in esports FIFA no such APIs are present therefore techniques such as machine learning and computer vision are required for the accurate detection and tracking of fast-moving football and players. Furthermore, the play-field has to be modeled with precision to enable the mapping of the ball and players within the play-field.

1.3 Research Questions

To achieve the aforementioned objective, the following research questions are answered in this study.

1. *How to accurately detect and track ball and players in an eSport FIFA 2022 game video?*
   Football is a very fast moving small object and it's very challenging to detect and track it precisely. Similar challenges are also encountered in player detection and tracking. Based on the literature review models will be developed to detect and track the ball, and the players on the football field. Afterward, the best-performing model/models will be used to extract the location in the form of the coordinate of the bounding boxes of players and the ball. These locations will be used in answering the following research question.

2. *How the extracted data can be used to determine the game tactics and strategies of the player?*

To answer this question, the data extracted in the form of coordinates from the above-mentioned model will be used. By using the ball and the players' locations, statistics such as who is in the possession of the ball, information related to the ball path on the football field while attacking for the goal, which part of the field is mostly used by the eSports player to take the ball to the goal, was he attacking or defensive during the match. To find changes in tactics of eSports players from attacking to defensive or vice versa over time, the statistics for every 10 minutes will also be calculated.

1.4 Challenges

Analysis of the visuals of eSport football matches can be a highly challenging task due to the continuous involvement of fast-moving objects such as football and players having simultaneous movements. For the match analysis, the extraction of ball and player motion features is of utmost importance for understanding the tactics and strategies of a team. The ball is a small and ultrafast moving object which can attend up to 270 km/hr. speed during the match. It can be easily occluded by the body parts of the players and in some instances, its color matches the jersey color of the players. Furthermore, Its color and shape also match certain play-field features such as penalty kick spot which makes it hard to be detected. In some cases, it is difficult to define ball possession if it is having close to equal distance between multiple players. Also, its shape can be distorted due to high-speed movement. Player detection and tracking share some of the ball detection challenges such as occlusion by other players, noise (from a color point of view) from the audience in the background, etc.

Typically in the eSport matches, a single camera that rotates at different angles is used. Due to this setup, the advantages of having the multi-camera system are not present which creates another set of challenges in the ball and player detection tracking process. Due to the camera movements, it is also difficult to map the play-field because this movement changes the size and pixels of the field. . This changing size makes mapping of the ball trajectory very difficult in subsequent frames because of variation of the origin coordinates of the field in subsequent frames. In the present research, the focus is on the investigation of these challenges and finding possible solutions to mitigate these issues while trying to detect and track the ball and players and the play-field modeling.

1.5 Thesis Structure

The present research focuses on the offline analysis of an eSport football match by detecting and tracking of ball and players from a match video. Since eSports analysis is a new research area, most of the existing studies refer to the analysis of a real football match which has several similarities with the eSport football match. Therefore, a literature search related to the analysis of actual football matches is presented in Chapter 2. In this literature review, potential methods for the detection and tracking of ball and players are also studied and identified. Chapter 3 describes the methodology used to answer the formulated research questions by developing models suitable for analyzing the match. In chapter 4, the implementation of the methodology is described.

This chapter is followed by results (chapter 5). The thesis is finalized with a discussion and conclusion and future outlook sections.

2. Literature review

The eSports games such as FIFA 2022 are generating interest nowadays. Therefore, the computer vision based analysis of FIFA 2022 football matches is also gaining momentum in the research community. The correct identification and prediction of the ball and players are of utmost importance for various reasons. The general public might want to watch the game from any arbitrary angle. For the coaches and players, the understanding of the tactical movement of players is important so that they can make their plans to counter the strategy of the opposite team. The sports broadcasting channels need to annotate, store, summarize and catalog the game based on the game's events so that they can easily retrieve a particular event of any game. This can improve the efficiency of their process and provide useful historic information about a game to the public. In this chapter, a detailed literature review related to the ball and player detection techniques based on deep learning methods and play-field modeling is presented in the following sections.

2.1 Ball detection and tracking

There are several sports in which a fast-moving ball is used and the discussion in this section is useful for all these sports. In the analysis of any game involving a fast-moving ball using machine learning (ML) techniques, the detection of the fast-moving ball is an important component that comes with several challenges. Table 1 summarizes features and challenges based on sports.

Table 1 Sports involving balls are categorized based on their important features and challenges [1]

Sport	Dominant features/advantages for computer vision based analysis	Challenges
Pool	Small region needed	Small ball size
Billiards	Ball color	Varying ball velocity
Table Tennis	Fewer cameras needed	Extremely high-speed and small ball size
Tennis	Court lines (Tennis)	Medium ball size
Golf	Little or no occlusion	High ball velocity
	Fewer cameras needed	
Volleyball	Bigger ball size	Same team occlusion

Sport	Dominant features/advantages for computer vision based analysis	Challenges
	Slow ball velocity	
Basketball	Bigger ball size in comparison to a tennis ball.	Same and opposite team occlusion
Football	Lines and circles on the ground	Players can possess the ball for a long time
		More cameras needed
Hockey	Lines on the ground (Baseball)	Medium ball size
Cricket		Occlusion
Baseball		High ball velocity
		More cameras needed
American Football	Lines on ground	Very heavy occlusion
Rugby		Spheroidal ball shape
		Irregular trajectories

As can be seen in table 1, the detection of a fast-moving ball in any game is quite challenging and comes with various technical challenges. The ball detection process involves several steps which can help in a clear identification and more accurate detection of the ball. These steps include positioning of the camera, modeling of the field, camera modeling and calibration, background extraction, and ball detection followed by ball tracking. In the next section, background information on all these steps will be provided.

2.1.1 Camera placement, model, and calibration

For an accurate and complete analysis of any ball game, we need to not only track the ball but also players, bats, rackets, etc. therefore camera placement, numbers, and camera movement is also important for accurate detection of the ball. In most of the investigations, a single fixed camera is placed for capturing the ball/player movements [2]. In some cases, the camera is placed facing the goal-face plane from one side of the football field [3] or the camera is placed to only focus on the goal area [4]. Typically, cameras having high speed, low-resolution, and high-vision are utilized in these methods. To cover a broader field even ceiling-mounted camera placement was also used in some cases [5]. For improving the view of the ball and tracking it from different angles, a two-camera system was also used in several investigations. It can be used in 3D localization of the ball/players etc. Several researchers have used stereo cameras for recording a ball game in which these cameras are synchronized [6], unsynchronized[7], or even dynamically coordinated [8]. In some cases, multi-camera systems are also used for coverage of the complete field. This type of system is more complex and expansive but it can drastically improve the 3D localization of the field. Also, every side of the field, ball, and game itself can be tracked easily. From this perspective, 4 and 6 camera systems were installed and used for tracking the ball [9] [10]. In summary, a single camera system is used when the field of view is fixed or quite small. This system has challenges in terms of occlusion of the ball by players which can be mitigated by using multi-camera systems. Furthermore, the multi-camera system can cover the almost complete field from different angles and improve detection accuracy. For a better 3D localization, a multi-camera system is useful and it can provide a lot of information about the ball and player movement which can help to understand the movements of the players thereby providing input in the game strategy.

The calibration of the camera is required to make sure that the final image provides the accurate and consistent mapping of the real-world image concerning the location, size, and graphical representation. In this process, the external and internal parameters of the camera are determined and calibrated for accuracy. These mainly include the ideal pinhole camera model [11], and the Tsai camera calibration model [12]. Figure 1 shows the basic process for the Tsai camera calibration model.

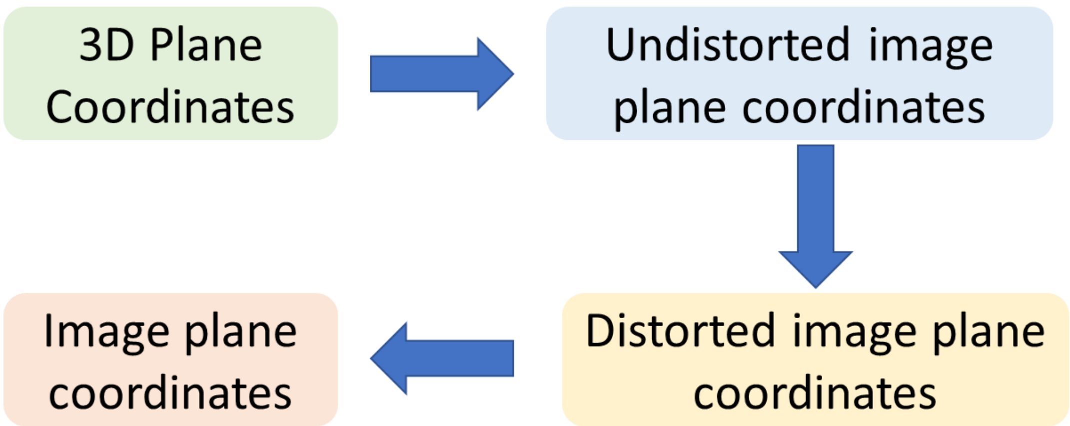


Figure 1 Basic process of Tsai camera calibration model

2.1.2 Background extraction

In a video, some pixels change their intensity drastically when they move from one frame to another and they might appear as moving even though their location is supposed to be the same. This creates trouble in locating balls or other objects with reasonable accuracy. These pixels have to be extracted for improving the ball detection and it is called foreground extraction. Similarly, some pixels do not change (or slightly change) their position in the background extraction. In this process, a trackable object such as a ball has to be removed from its background then only accurate detection and tracking are possible. In the present case, a ball is a target and in the background players, field, markings on the fields, spectators, and ground staff all are also part of the image//video. Therefore, a lot of noise comes from these background objects while tracking the ball.

Odobez and Barnard suggested using color-based patterns for recognition of the playfield [13]. They used a Gaussian Mixture Model (GMM) for modeling the color. This model is then further adapted using a Maximum a Posteriori (MAP) adaptation [14]. Afterward, they trained the model using several grass images from several football fields. The trained model is tested for various fields used in rugby, hockey, and football, and this model is adapted by using the MAP adaptation technique for different fields. In some cases color only cannot provide the required results and geometry-based masking methods are also used for foreground extraction[15], [16]. In some cases, color and geometry-based masked are fused for this purpose [17].

2.1.3 Ball detection

Various ML techniques have been utilized for the detection of the ball in different ball games. In the earlier research, mostly probability density and image edge features-based algorithms such as Meanshift, Particle Filter, and Kalman Filter, were used as tracking standards. In the Meanshift method, the object to be tracked such as a ball is searched following the path of increasing probability gradient and iteratively converged to the local highest point of the probability density distribution [18]. The ball is first modeled using some feature such as color distribution and its probability distribution is calculated in the subsequent frame. This method is mainly suitable for tracking balls when the background is having a different color than the ball.

In the particle filter method, the ball to be tracked is modeled and some particles are interspersed in the space [19]. Afterward, a similarity matrix is defined which is used for finding the match level between the ball and the particles. When the ball is searched, the method produces some more particles following a predefined distribution such as Gaussian distribution. The similarity between the ball and particles is counted which provides the position of the ball. Then, in the next frame, more particles are spread to increase the tracking accuracy.

The Kalman filter takes a different approach by modeling the ball’s motion rather than its features [20]. It is mainly used for tracking the ball in the next frame. Another similar method is called optical flow tracking. In this method, the ball’s feature elements are retrieved and then the optical flow tracking points are determined in the subsequent frame. The location of the ball is determined using statistics. These classical methods had several limitations such as not being able to handle complex tracking challenges, and lower robustness and accuracy as compared to the new advent of deep learning approaches.

Kamble et al. [21] used a novel deep learning approach in which an object is classified into 3 classes ball, player, and background, for two-dimension football detection and tracking from football match videos. They proposed a two-stage buffer median filtering background modeling for the detection of any moving object. For validation of the ball tracking, they used the probabilistic bounding box overlapping technique. Furthermore, they developed and used full and boundary grid concepts for resumption of any lost tracking of the ball in case of loss of ball track or ball out of frame situations. The proposed approach showed tracking of the moving ball with high accuracy and robustness even in the case of small ball size and fast movements.

Huang et al. [22] analyzed tennis matches in which tracking an extremely fast-moving and small tennis ball was a big challenge. The correct mapping of ball trajectory is important in analyzing the performance of the players and can provide insight into possible improvements. They developed a deep learning method “TrackNet” for tracking the tennis ball from ordinary videos of a tennis match. It is indeed a challenging task because many times ball image is too small, blurry, or even invisible (see figure 2a). To overcome this challenge, they proposed to use a deep learning method based on heatmap which is trained not only to spot the ball in each frame but also to learn the trajectory of the ball from consecutive frames. It consists of both convolutional and deconvolutional neural networks. A total of nine videos were annotated and used for training purposes. They used men’s single final video from Summer Universiade (2017) for analysis purposes. Using this method, they were able to precisely determine the position of the high-speed tennis ball and their method provided precision, recall, and F1 measure reached 99.7 %, 97.3 %, and 98.5 % respectively.

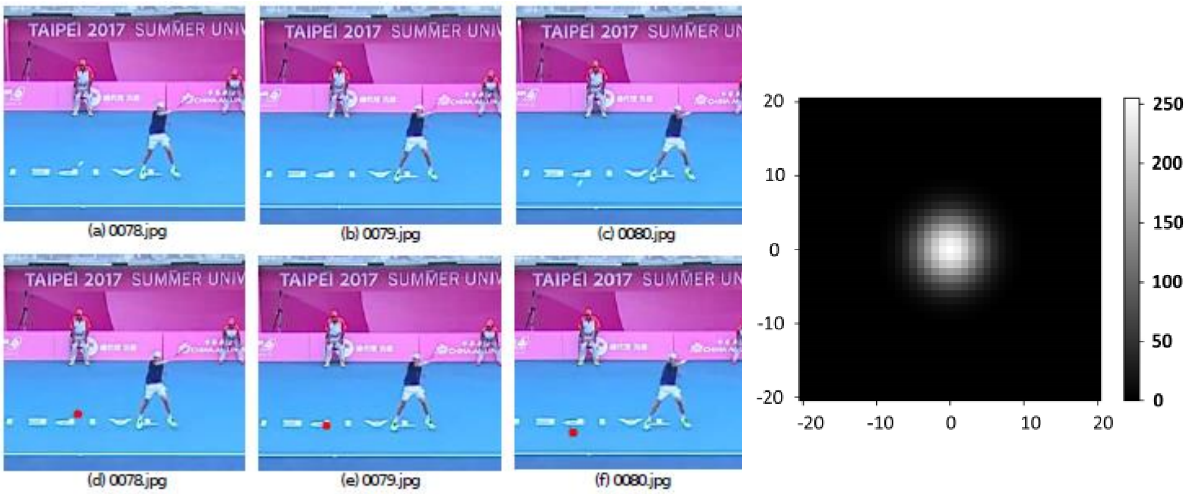


Figure 2 (a) The ball image is hardly visible in several frames (b) Detection heat map of the tennis ball [22]

Similarly, moving ball detection is important for understanding the action recognition and prediction in a handball match. Buric et al. [23] compared the performance of two convolutional neural network-based methods for the detection of a moving ball in a handball match video. They compared the speed and accuracy of the two methods and also compared the impact of training on the final detection results of the ball. Jiao et al. [24] provided a detailed overview of methods covering one-stage and two-stage detection processes that can be utilized for object detection. The one-stage detector e.g. YOLO [25] and SSD [26] are relatively simpler in use and provide high inference speed. The 2 stage detectors have high object detection accuracy in comparison to single-stage detectors [27]. The two-stage process includes the first use of the region of interest in which the part of the image which is of interest can be extracted from the image and only this part is used for the feature extraction by the ROI Pooling operation. In a single-stage process, the region of interest operation is not performed instead, the prediction boxes are directly used for feature extraction purposes. This reduces the operation time and makes this method highly suitable for real-time devices. A depiction of the architecture of both of these processes is provided in figure 3. Since the target of current research includes real-time detection of a fast-moving football, a single-stage detection process (YOLO) could be utilized for detection purposes.

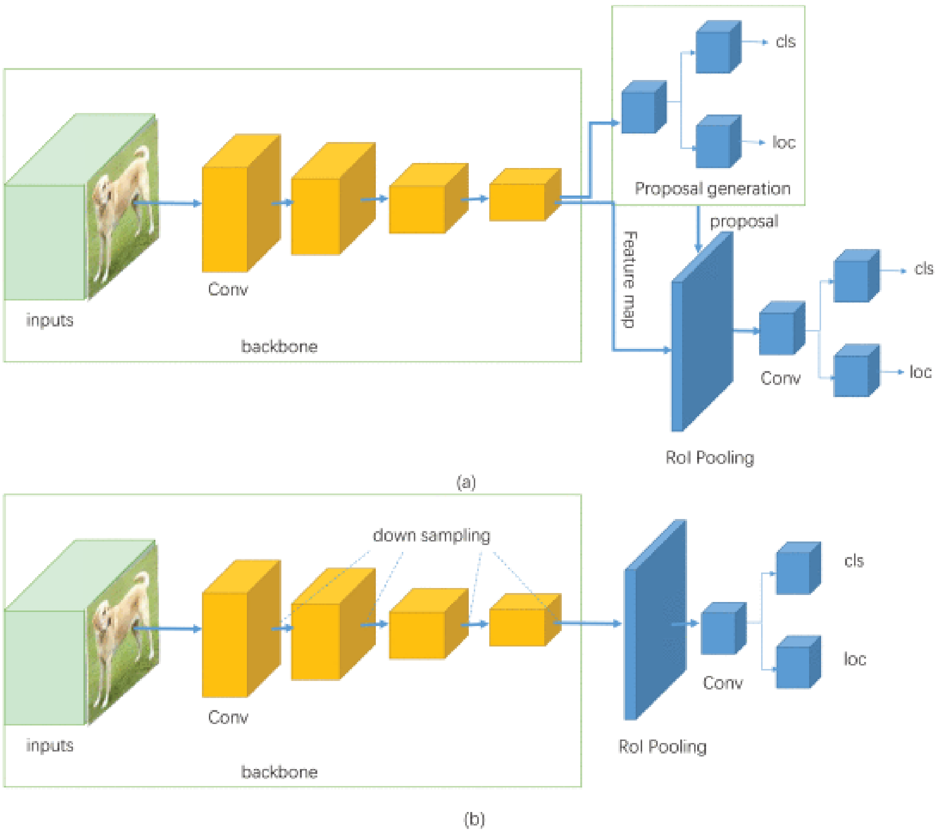


Figure 3 a) Basic process of a two-stage detector which includes region proposal network b) Basic process of single-stage detectors [24]

The one-stage detectors utilize prediction boxes from images or frames of a video The majority of the object detection method use deep learning-based methods for extraction of features from the videos/images [28]–[31]. Chakradhar et al. [32] used deep learning-based scoreboard detection for automated extraction of key events utilizing the YOLO method. They used a database of 1300 images for training their detection algorithm YOLO. They detected the scoreboard from the video and processed the scoreboard part of the image for the reduction of false positives and noise reduction. These processed images are passed through an optical character recognizer for obtaining the score of the match and further processed these images using a rule-based algorithm for the generation of events timestamps. They used this method for cricket ball image detection in a cricket match.

Often tracking a fast-moving object such as a ball comes with several challenges including a blurred image of the object in the video and the ball position does not coincide in the successive frames of the video. Typically these issues are tackled by background subtraction and utilization of deblurring algorithms [33] [34], [35]. Zita and Sroubek [36] proposed an innovative tracking by segmentation approach in which a possible fast-moving object sequence producer was implemented for training pipeline and demonstration of different fast-moving object detection scenarios. Their architecture was based on a deep network and ENet which is a U-Net architecture, is the main pillar of the proposed method. The main idea behind the segmentation approach was to enable the tracking of fast-moving objects with white color and having no noticeable texture such as ping pong, squash, badminton, and tennis. Due to this color challenge, ball detection in the aforementioned sports is difficult and cannot be achieved by using a single image segmentation approach because the detection method falsely recognizes any bright color in the image and segment the image falsely based on a bright spot/line detected on the image. To mitigate this issue, the authors used sequencing of several subsequent frames as a network input which improves the trace uniformity in time. This method helps in overcoming the existing limitations such as longer computational time and false detection of fast-moving objects.

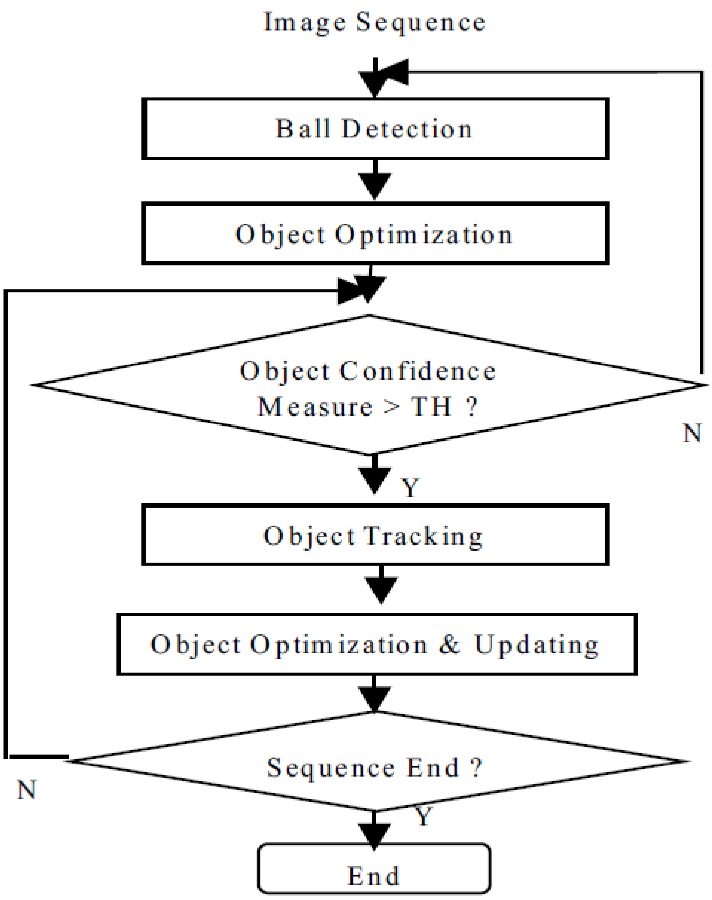


Figure 4 Framework of the proposed algorithm by Tong et al. [37]

Liang et al. [38] detected and tracked a soccer ball in a broadcast video. From several consecutive frames, the soccer ball is identified based on its color, shape, and size. Afterward, a weighted graph is developed with each intersection depicting a candidate and each edge linking 2 candidates in adjoining frames. They used the Viterbi algorithm to calculate the optimum path for finding the ball's location. They used Kalman's filter-based model matching to follow the ball in successive frames [39]. This method increases the vigor of the ball detection and provided promising results even in case of bad field (video) conditions. Tong et al. [37] proposed a different way of ball detection and tracing in a real football match. They used an indirect detection method in which first they identified and extracted the football field and all the actions were restricted within this field. First, the areas without balls were removed from the identified field based on ball color and shape. Afterward, the leftover areas of the field are examined from a finer perspective and an optimum one is identified as a ball. The framework of the proposed algorithm is shown in figure 4. For the ball detection, shape, size, and texture features were selected and then Kalman filtering and template matching were used for ball tracking. Hough transform-based circle detection algorithm was used for soccer ball detection [40]. Although for certain conditions the proposed approach worked well there are other conditions such as ball size is too small or ball visibility is hindered by the player or background field is too complex which the proposed approach failed to provide satisfactory results.

Another set of methods for object detection from images/videos is template matching which is gaining a lot of traction despite having some shortcomings. Broadly speaking in template matching, a small part of an image is found by matching a template. There are several variants of this technique such as Matched Spatial Filters, Synthetic Discriminant Functions, Principal Components Projections, and Reconstruction Residuals [41]. Kumar et al. used Spatio-temporal template matching for the detection of a basketball, the recorded game had several cameras for the capturing game from different angles [42]. First, they defined a 3D grid around the basket and then try to detect the possible candidates for the ball by finding the match of the ball template in each frame. Afterward, the false positives are taken out using the integration of the sequential assessment of the ball path with the Random Sample Consensus model. Figure 5 shows an example result of their analysis. Their method showed promising results with an 80 % correct and 4 % false detection rate.

Figure 5 a) The reality in blue and ball candidates in orange b) Ballistic (green) and non-ballistic (orange) ball candidates.

2.2 Player detection and tracking

In popular sports such as football, field hockey, American football, and cricket, player detection is also another important component of sports video analysis. From a comparison perspective, the pedestrian detection framework is quite similar to the player detection framework. In the pedestrian detection framework, a person who is walking or running on the street is detected for several purposes. This detection can be useful for automobile applications for safety issues. It is also used for surveillance by government agencies for potential criminal activities. The player detection framework depends upon the game as the movement of the players varies a lot because of the nature of the game.

In many games, several players are wearing the same type of uniform and trying to achieve the same target. Their movements, directions, and locations could be quite varied and spread across the field. These movements are usually captured by cameras which are placed at a certain distance and the resolution of the player image taken could be below which can create trouble in detection. Therefore, certain player detections algorithms are used for handling variations such as low resolution, and uniform color. occlusion, weather conditions, illumination, etc. Therefore, the detection process has a certain level of challenges which makes it different than the pedestrian detection framework. Typically player detection can be performed on a static image and afterward, results can be optimized by utilization of subsequent images from a video. Alternatively, multiple frames can also be captured from different angles using different cameras and then optimized with multiple frame analysis.

Typically playground is homogeneous in color and mostly it is green in color as the color of grass. Many researchers have used this feature for performing player detection in a match. Several studies used broadcasting videos and use one color which is dominating feature in the detection of players[43][44] [45]. The same method was also applied to the detection of basketball players by using HSV color histograms by which the main color of the playground was determined [46][47]. In these studies, first, the leading ground color is determined by the histograms of HSV and then the ground is segmented to detect the players. In another study, the same process of the color histogram was used but instead of using HSV-based histograms, they utilized RGB bases color histograms and eventually identified the playground in the videos[44]. In another study, the Gaussian method in RGB color space was used for finding the color classes and performing region segmentation [48].

For successful detection and tracking of players, the correct positioning of the camera is also important. There were several studies in which different camera configurations were used. In previous investigations, a multi-camera system was used for this purpose by [49], [50]. In this configuration, the main objective was to capture the player movements simultaneously by using multiple static cameras. For the identification of players, the background subtraction method which constantly updates the model because of changes in conditions such as shadows, objects, and background lights, is used by several researchers [49], [51]. Other studies used statistical methods which are more useful for simpler cases [52]. In some instances, a camera is placed roughly at the center of the field having a wide angle which makes it possible to view all the players at the same time[53]. This makes it possible to track all the players together which can help in understanding their simultaneous movement. This position provides a low-resolution image with smaller heights of players which makes it difficult to detect and track the players. Figure 6 shows such a type of arrangement.

Figure 6 A sample image from a wide-angle view camera from a football match[53].

Another approach is to use perform a framewise detection. In this approach, the subsequent frames are then linked together for continuous detection. The main limitation of this approach is to have the dependency on the precision of the object sensor. Hurault et al. developed a novel approach in which players were not manually annotated but were detected by a self-supervised learning approach and the association between the consecutive detections were established [53]. This method provided competitive results in the detection mechanism concerning the conventional methods but it has shown some shortcomings while tracking the players.

In another study, multi-object detection and tracking are performed using a support vector machine and particle filter technique [54]. They first developed an improved particle filter (SVR particle) for tracking the players. For this purpose, SVR is integrated with the Monte Carlo framework and eventually enhances the player tracking efficiency. A typical framework for a multi-player tracking algorithm is provided in figure 7. In this study, SV classification is used for automatic detection and it is combined with the playground segmentation for player extraction from the region of interest. This study led to a unified framework for automatic detection and tracking of players.

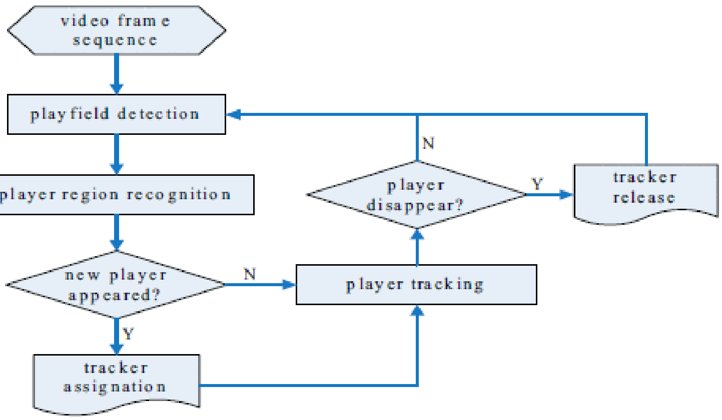


Figure 7 A framework for multi-player detection and tracking [54]

Yang et al. proposed an enhanced particle filtering for player detection and tracking [55]. They used salient region detection for ground segmentation and afterward detected the football players by using the edge algorithm combined with the Otsu algorithm. Figure 8, shows the results of segmentation using salient region detection. First, the salient values from image pixels (left) were used to compute the salient maps (middle). Then, salient regions were segmented (right) from salient maps. They also combined color and edge features for detection purposes. Together with an enhanced particle filter, the proposed algorithm provided better accuracy results. It was able to quickly and accurately detect and track the players in a football match.

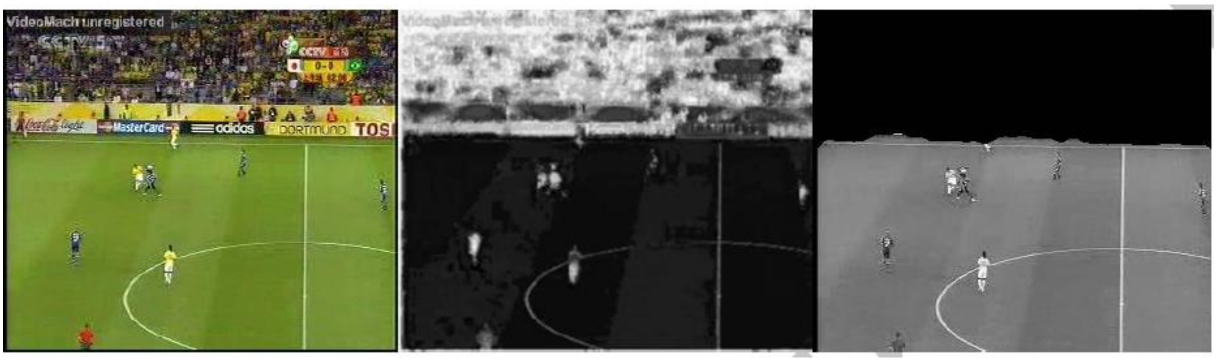


Figure 8 The segmentation result by salient region detection (left – original image, middle – HC map, right – segmentation result)

Zhang et al. compared several models for player detection and concluded that classic algorithms end up losing the player body due to overlapping and blurred images of the lens because of faster movement [56]. To mitigate this issue, they integrated YOLO [57]and Deep Sort [56] which are 2 main algorithms for the detection and tracking of objects respectively. Their results were satisfactory and they were able to detect overlapping player bodies. Naik and Hashmi proposed a deep learning-based YOLO model (v3) for the detection of football and players in a broadcasted football match video. They first cleaned the video by removing the non-relevant parts such as replays, zoom-in, breaks, etc. from it. They tracked the players by employing the SORT algorithm with Kalman filter and bounding box overlap. The YOLO model is a completely trained neural network model which can find objects in an image by using the bounding box location. Figure 9 shows a schematic diagram depicting various blocks of YOLOv3 architecture. It uses logistic regression to estimate the extent of overlap of the bounding box based on its comparison with the previous such value. The estimate is not considered if it is lower than the threshold. It also uses three scales for predicting feature extraction.

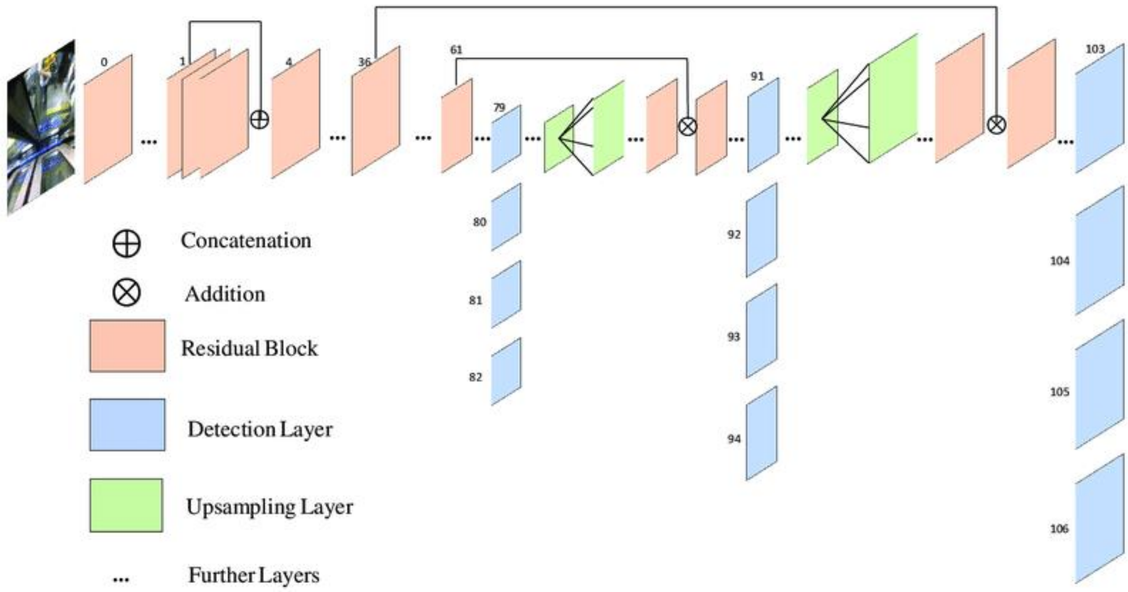


Figure 9 YOLO V3 architecture showing blocks and associated number of layers in the blocks and their functionality [58].

In this research, a significantly big-size data set was used which removed the data augmentation requirement. Figure 10 shows the various steps used in the process of multi-object detection in the proposed model. They performed a comparison of their results with the existing solutions and also carried out a performance analysis of the model results. The model was able to detect the ball and players despite having occlusions on several occasions.

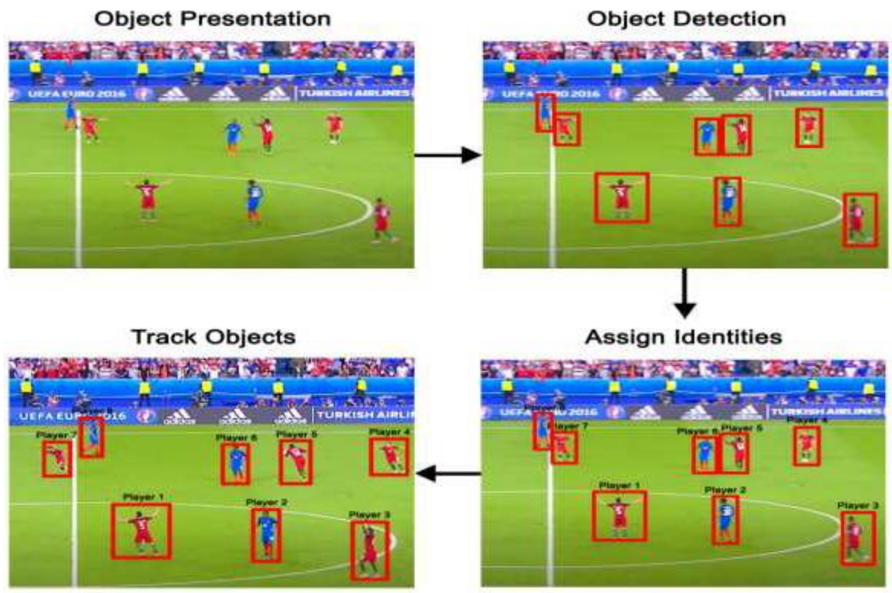


Figure 10 Multi-object detection process proposed and used by Naik and Hashmi [58]

Yoon et al. used deep learning-based techniques for the assessment of a Basketball match from videos having inadequate conditions such as dynamic shifting of camera position etc. [59]. Through these techniques, they were able to find the ball pass relationships between the players. They used YOLO for real-time detection of the ball and players. It uses Darknet which involves convolution neural networks to classify the object detection. For improving the performance of the proposed method even in inadequate conditions, they used location coordinates from the previous or subsequent frames. The end-to-end process used by the authors is given in figure 11. First, the video frame is dissected in a grid followed by the putting bounding box estimation. The bounding box has a width, height, and a central location. This technique uses K-means clustering to find different groups from the collected ground truths. It is used to find the score based on the class and confidence. Although the proposed approach showed some shortcomings and inaccuracy due to inherent limitations in deep learning approaches, still the proposed method provided valuable insight into the game itself. This technique is faster than Fast-RCCN and Faster-RCCN.

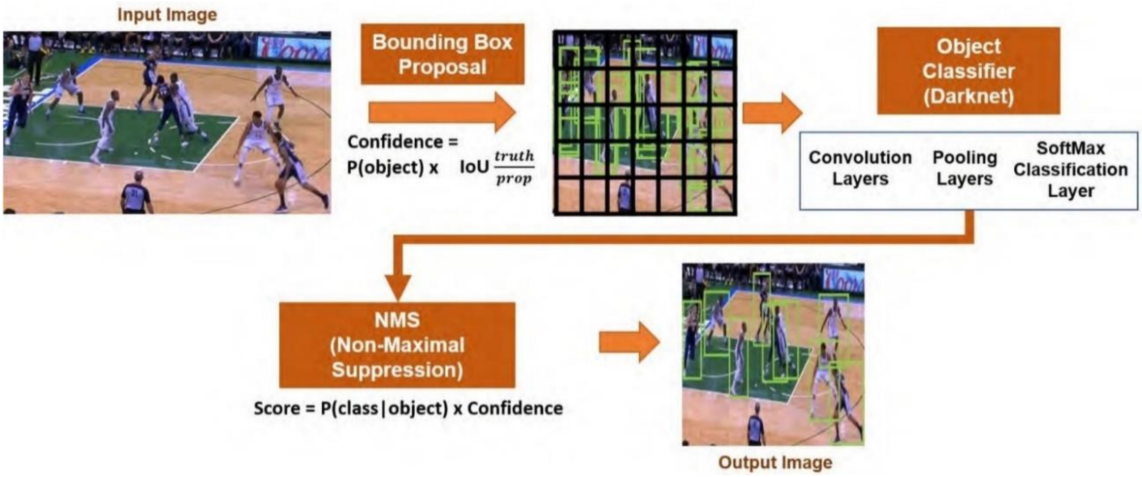


Figure 11 The process of object detection and classification using YOLO [59]

Alahi et al. proposed a generic approach for the detection and tracking of people with stationery and one-directional cameras which can also have degraded images because of the shadow effect [60]. They solved a sparsity-constrained inverse problem in which a weighted scheme is considered for sparsity prior. They used this generic approach for the detection and tracking of players from a basketball game. They were able to successfully detect and track players despite having highly noisy features. This approach can be used for any type and number of cameras.

Liu et al. developed a method that carried out automatic multi-player detection, without supervision labeling and fast-tracking of the players in a football match video. They integrated the prevailing color-based background deduction and an enhancing detector. The enhancing detector used Haar features. They taught the appearance model by collecting player samples using the detector which provided the capability of detecting and tracking different videos without involving any manual loading. The implemented player detection process is shown in figure 12. They utilized Monte Carlo for player tracking which was improved by using data-driven dynamics. This method is implemented in FIFA WC 2006 videos and was able to provide good detection and reliability of tracking of the players.

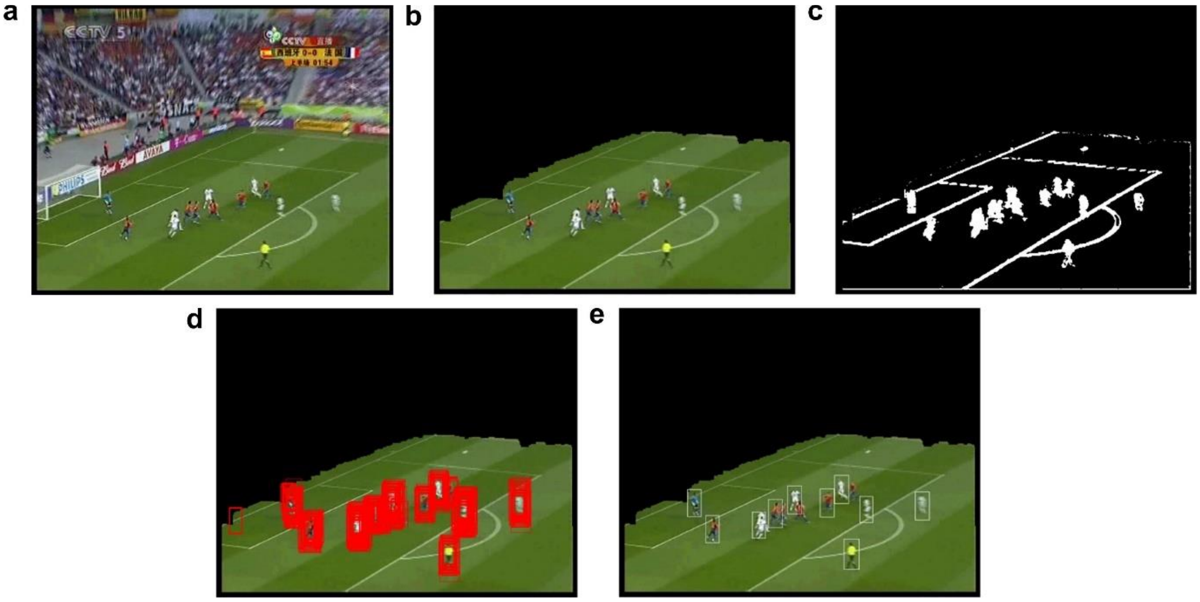


Figure 12 Player detection process a) Original image b) playfield filter out image c) player mask d) initial response of player detector e) final result after post-processing [47]

Buric et al. assessed a handball match video having a single source of the video [61]. In the handball match, the players move in and out of the video angle rather more frequently, change direction quite frequently, and provide occlusion considerably, which makes the tracking a formidable task. Furthermore, players have similar sports attire, and the background color changes frequently within the field due to the use of several colors while painting the handball court. They proposed the use of tracking by detection methods such as CNN for tracking the player location in a handball match, due to their higher accuracy, low computational requirement, and the possibility of doing it online. They used three separate tracking detection methods; 1) the standard Hungarian assignment algorithm, 2) the simple online, and 3) the Real-time tracking (SORT) algorithm. Their results were compared to understand the impact of various methods on accuracy. The results from SORT algorithm showed improvement in player detection under poor conditions by using deep learning-based detectors that use the tracking by detection approach. Cases in which visual features are used to again identify players provided promising results. All the methods had issues in following the players who are temporarily out of sight, thus player re-identification is still a challenge in this approach (figure 13). Still, the proposed approach showed promising results and a further improvement in determining the trajectory of the players could be an interesting addition to the proposed method.

Figure 13 The tracking situation example for Hungarian problem (top), SORT (middle), and Deep SORT (bottom). The left and right frames have a difference of 1 second.

Sun and Liu used a template matching method for the detection of players in a football video [62]. They first segmented the players from the frame of the viewing angle and then the template matching is performed for player detection. They matched 4 types of templates and their method was able to detect players with > 90 % precision.

2.3 Play-field modeling

For the play-field modeling, one of the most used features is the color of the play-field as it is usually homogenous thereby making it an easy-to-use feature. In previous efforts, rule-based leading color models were used [63]–[65]. These models are flexible in accommodating different colors by just changing the color subset by changing the range. The play-field color has also been represented using a Gaussian mixture by Luo et al. [64]. In other studies, unsupervised learning approaches were used for detecting the play-ground color [63], [66], [67]. Similar to the field color, the line marks were also used for the field modeling. Field color can support finding the line marks reliably.

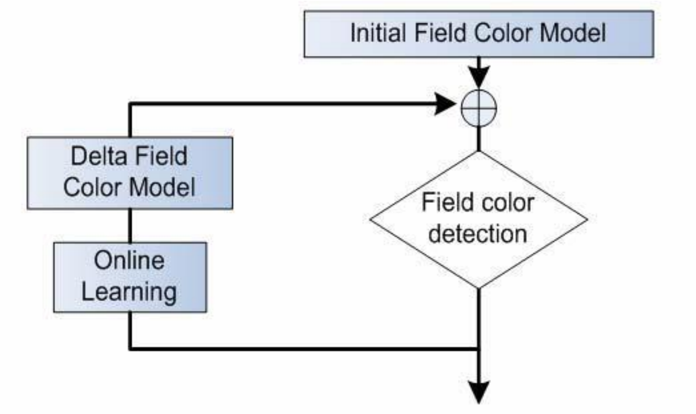


Figure 14 The scheme of the hybrid color model

Liu et al. proposed to use a hybrid color model consisting of the initial and delta models [68]. The model scheme is shown in figure 14. In this hybrid model, first, an initial model is developed which used offline data for training purposes. The pixels from this data set were labeled to estimate the pixel with field color and without field color. They assumed that the field color can have only three ranges of color to ease out the implementation process. They expected that only shadowing and lightning can affect the field color distribution. This method is used to develop a labeling tool in which three color ranges can be defined in the HSV and the rest of the space was assumed to be a non-field color. The result is shown in figure 15. The knowledge gained during the literature review is used in finding suitable methods for the analysis of the FIFA 2022 games. In the next chapter, an in-depth review of the aforementioned three methods is carried out and further explanations are provided.

Figure 15 Play-field color labeling tool [68]

3. Comprehensive review of the proposed methods

The detection and tracking of ball and players are one of the main challenges in analyzing a FIFA football match. As explained in the previous chapter, several methods can be applied to this process. The most suitable methods related to the ball and player detection are identified in the previous chapter. These methods are comprehensively reviewed and described in this chapter. Furthermore, a method is also selected for mapping the play-field which is necessary for any successful analysis of the game. This method is also explained in this chapter.

3.1 Ball & Player detection and Tracking

Although ball detection is a single object detection problem unlike soccer player detection which is multiple object detection, still detection of a ball is quite more challenging because of its high speed and sudden change in motion. Apart from this, there are other challenges while detecting the ball such as its size, being a small object in every image makes its blob size smaller and therefore it is difficult to derive and characterized features from it. Occlusion and overlapping with players are additional challenges that needed to be dealt with for the continuous detection of soccer ball[69]. Whereas player detection will be based on team-based detection.

As a ball and player detection in each frame is an important requirement to form a consistent trajectory. This trajectory should be able to keep a record of the ball and player's location from the first frame should be able to predict the correct location of the ball in case of occlusion and if two or more ball candidates are available then it should be able to predict the correct ball candidate. In this section methods that can be used to detect and track ball and the players have been discussed.

3.1.1 Gaussian Blur

Image blurring can be done by convolving the image with a low pass filter kernel. Gaussian blur is very useful in removing Gaussian noise from the image, although it does not preserve the edges in the input. In this process, the pixels near the center of the kernel are provided with more weight than those far away from the center. This averaging is done on the channel by channel basis and later this average channel value will become the new value of the filtered image. Figure 16 shows an example of a two-dimensional Gaussian filter [70].

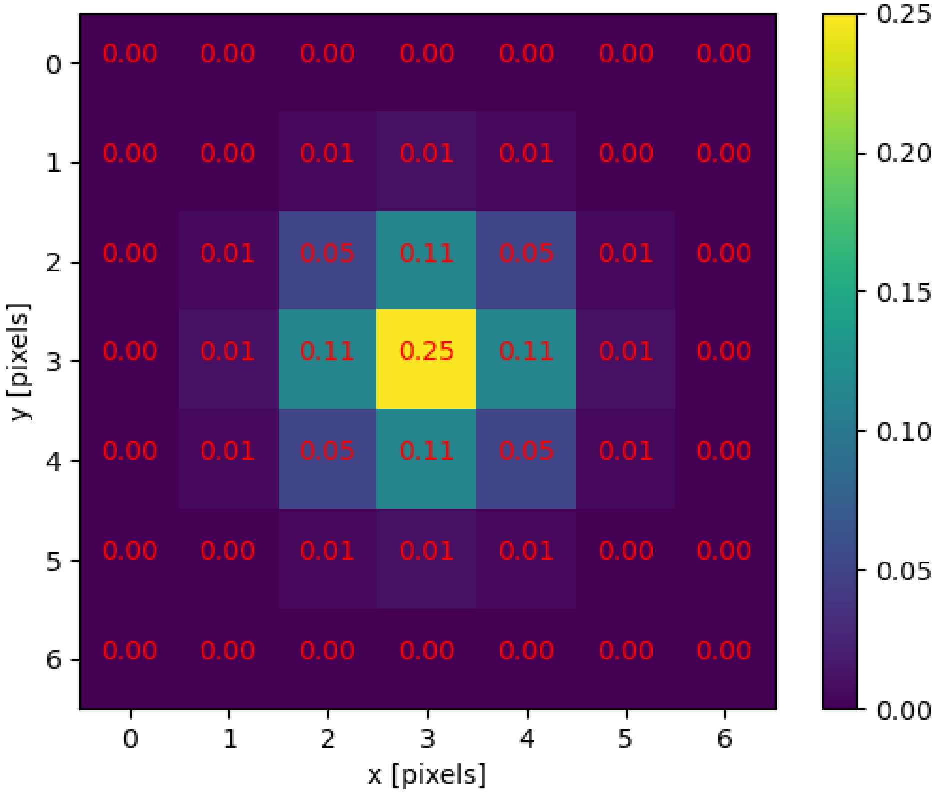


Figure 16 Two-dimensional Gaussian filter, showing pixels near the center with extra weight

Consider that this plot is laid over the kernel of the gaussian blur filter and the height of the plot in accordance with the weight given to the pixel in the kernel i.e the pixel at the center will become more important than the pixel of the outer side kernel. Mathematically Gaussian filter can be represented by equation 1

$$G_\sigma = \frac{1}{2\pi\sigma^2} e^{-(\frac{x^2+y^2}{2\sigma^2})} \qquad (1)$$

Where σ represents the standard deviation while x and y represent the distance of the pixel from the center of the image (i.e. top left corner)

3.1.2 Convert frame to HSV

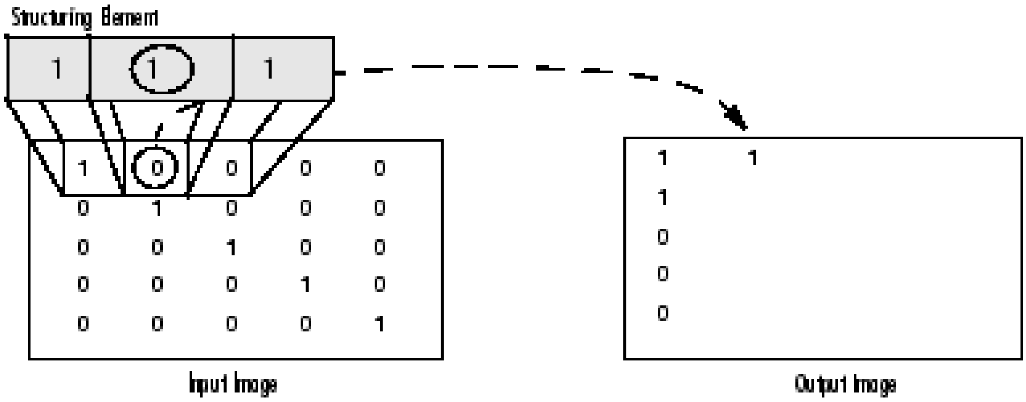
The next step is to convert the image from the RGB channel to HSV. This conversion is necessary to separate image luminance from color information, as in RGB separating color information from luminance is not possible. HSV consist of 3 matrices of “Hue”, “Saturation” and “Value”. The use of HSV rather than RGB space works better in texture analysis. Also being a perceptually uniform space HSV outperforms nonuniform RGB[2]. Not only HSV could be a better color space concerning RGB in the situation of color texture analysis, but it also displayed superiority in noise conditions as well.

3.1.3 Color thresholding and Binary mask

For foreground extraction to track the moving object(in this project ball and player) conversion of the color image to a binary image is required. And for doing this, thresholding the image based on the “Hue”, “Saturation” and “Value” of each pixel should be done [71].

3.1.4 Morphological operations

The resultant binary image might consist of many imperfections and distortions. To deal with this problem morphological operations are performed. Usually, the morphological operation is a two-step process of erosion and dilation.



. Figure 17 represents Morphological Dilation over a Binary image [72]

Morphological operations are performed on the binary image for operations such as segmentation, edge detection, enhancement, etc. Dilation adds the pixel to the boundaries of the objects in the image while erosion removes pixels from the boundaries of the object (Figure 17). And this addition and removal of the pixel are determined by the size and shape of the structuring element used in the processing of the image. This structuring element is a small 2D array of pixels that are used to operate over the image and then the image gets modified accordingly.

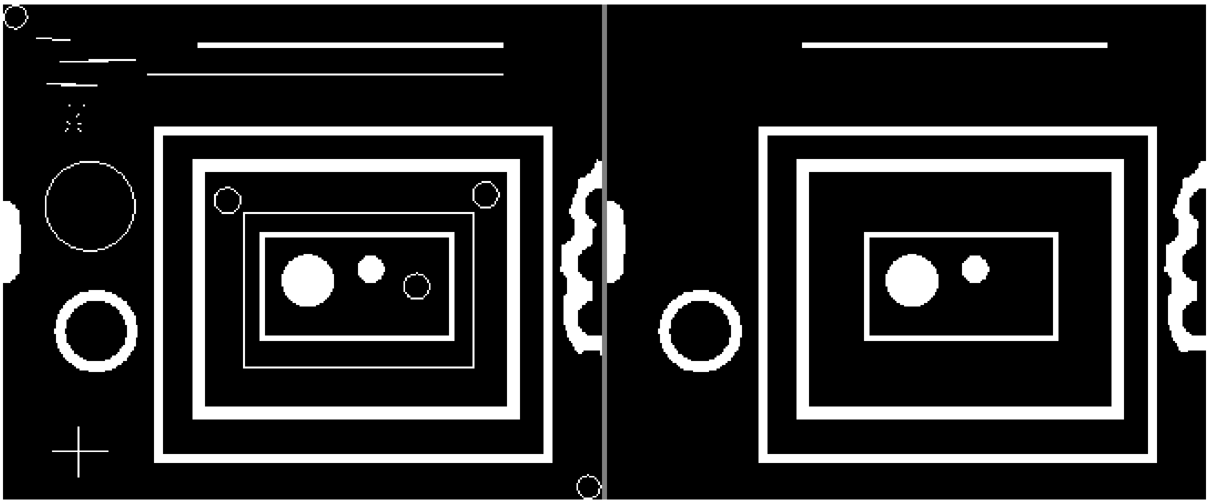


Figure 18 The image shows a morphological opening

Generally, the process of erosion and dilation are performed consecutively. When the erosion is performed before the dilation then this process is called morphological opening as this opens gaps among the objects of the image that are connected loosely and hence removes the small holes and small lines while preserving the shape and size of larger objects in the image as shown in Figure 18. And when dilation is performed first then it is called morphological closing as it closes the gap among the objects which are having weak connections while maintaining the shape and size of the comparatively larger objects in the image as shown in Figure 19. In our application of soccer ball detection, the morphological closing operation seems to be a better option as it will help eliminate the small discontinuity.

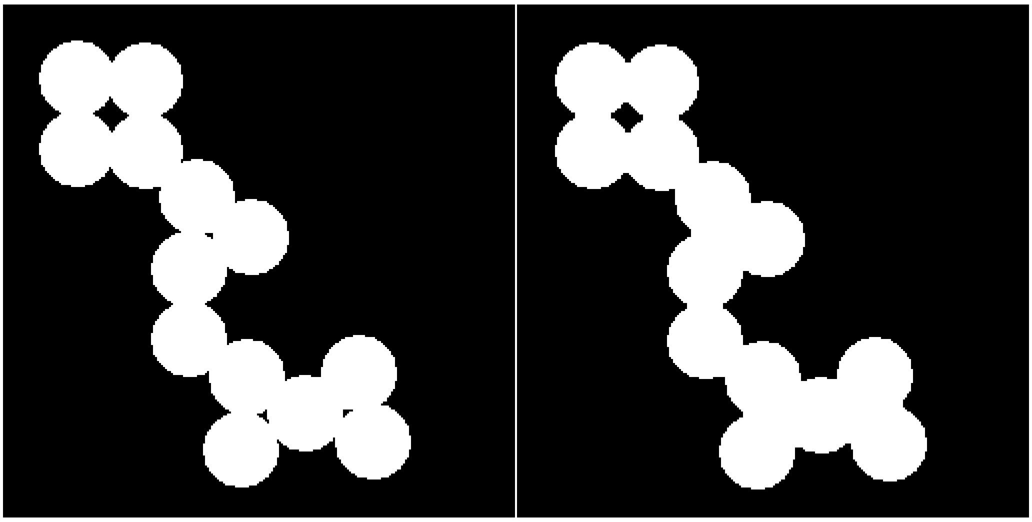


Figure 19 Image shows a morphological closing

3.1.5 Contour detection

A solid object, captured in the image is bounded by image curves, these curves are also called as contour or outline of the solid objects. Contour detection is the process of the curve joining all the continuous points which have either the same color or intensity [73]. It is one of the most useful tools for shape analysis and object detection. It's more convenient to find contours in binary images so the first step toward finding contours is to convert the image to a binary image either through thresholding or by canny edge detection technique. The binary image consists of 1’s and 0’s, where 1s are the location where contours can be found whereas 0’s are the background of the image. In this project contour detection algorithm by Suzuki et al. [74] can be used. This algorithm takes the image as an input and returns the matrix of the coordinates of the border of the contours.

These contours of the objects can be used to determine the area of the objects in the image, by calculating the enclosed area inside the contours [74]. Based on the size of the contour it is possible to differentiate between players and the ball. As the ball, players and noise are the only moving objects in the image, it is possible to distinguish between the ball and players by applying a threshold value of the size. Although this threshold value cannot be global and it needed to be set every time on different videos. Furthermore, this threshold value will be affected by the pixel size of the image as the camera will move and also be zooming in and zooming out.

3.1.6 Kalman filter

The Kalman filter technique is used to estimate the state of a linear system where the state is assumed to be distributed by a Gaussian [75]. It is a set of mathematical equations that generates a logical computational means to estimate past, present, and even future states even though not have enough knowledge of the modeled system. It is a feedback-based system. Kalman equation falls into two categories i.e. time-based equation and measurement update equations. Where time-based equations can also be considered predictor equations, while the measurement update equations can be considered corrector equations. Figure 20 shows the resemblance of the final estimator algorithm to the prediction correction algorithm for solving numerical problems.

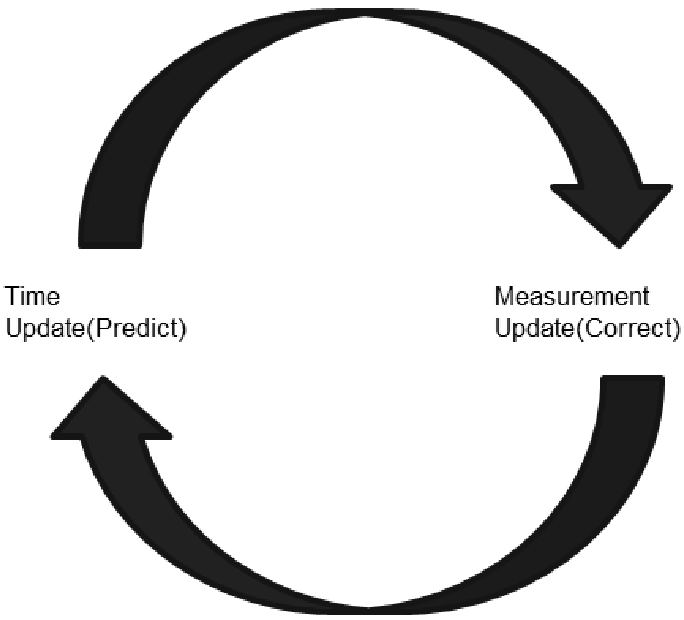


Figure 20 Discrete Kalman filter cycle

Although this filter works great on linear systems with Gaussian distribution, it performs poorly in situations such as occlusion due to missing input values [76]. Its performance is also not satisfactory in the case of non-linear motion. To overcome this occlusion problem such as in the case of soccer ball tracking Kim et al. [77] have used a Dynamic Kalman filter. According to their approach for every instance, they increase the search area, covariance matrix, and target to be tracked depending upon the situation. For example in the case of soccer ball occlusion the player who was in the possession of the ball just before the occlusion will become the tracking target and the search area will be shifted to the area near the player. Figure 21 shows the approach used by Kim et al. [77] This approach performs better than the usual Kalman filter approach. In this project as well we will be using the Dynamic Kalman filter approach.

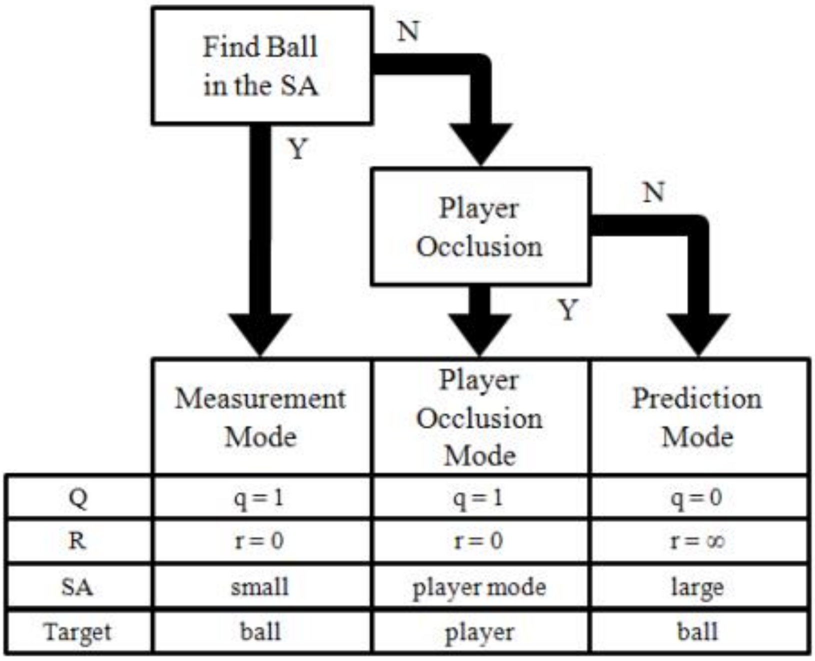


Figure 21 Dynamic Kalman Filter [77]

3.1.7 Template matching

Template matching is one of the methods of image processing where a base template image is provided to find the exact match in the source image. This can be done by comparing one pixel at a time in the source image with the template, and the output is received in the form of the array which has a similar value as the template image [78]. Edge detection and object detection are some of the most popular applications of template matching. Area-based approach and feature-based approach are the two main approaches to template matching and they can be used according to the problem and its application.

The area-based approach is also called the correlation-like method which is also a combination of feature detection and feature matching [78]. This method is useful with images that have weak features and also operate on the bulk of values. Where else in feature-based matching both source image and template image have more similarity in terms of features and control points such as curves or a surface model that has to be matched.

3.1.8 YOLO (You Only Look Once)

Yolo is a single end-to-end trained convolutional network that is capable of predicting multiple bounding boxes and the class probability of the boxes in an image. Complete yolo architecture is shown in figure 22. It trains on the full images to directly optimizes detection performance [79],[80].In this model, the input frame is divided into a small number of grids(S X S grids)Where each of these grids on the input image predicts bounding boxes for an object, along with its height, width, and x, y coordinates of the box, confidence score and conditional class probability. An example of a bounding box with its detection is shown in figure 23.

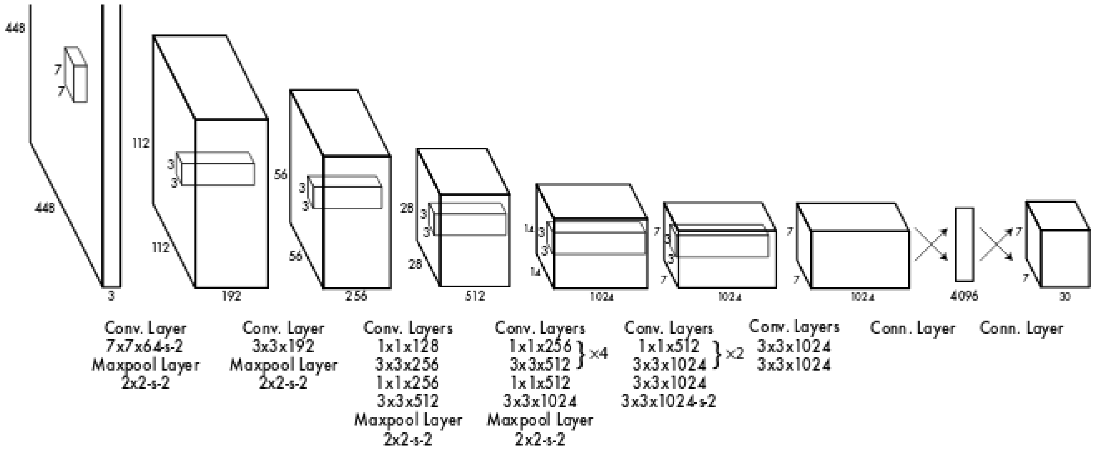


Figure 22 The fundamental architecture of YOLOv1 with 24 convolutional layers followed by 2 fully connected layers [80]

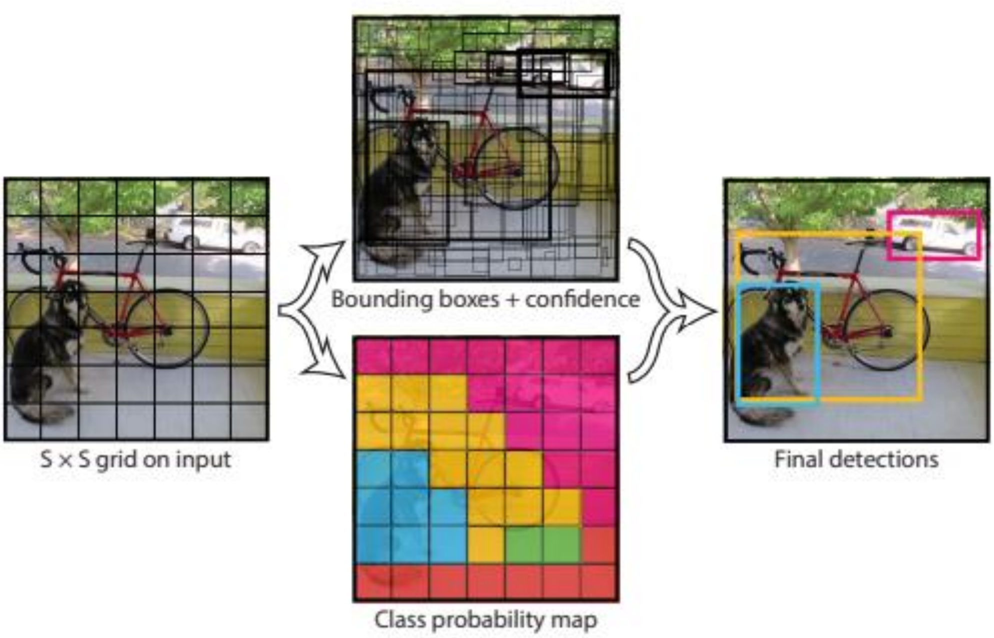


Figure 23 Yolo models detection as a regression problem. It divides the image into an S × S grid and each grid cell predicts B bounding boxes, confidence for those boxes, and C class probabilities [79].

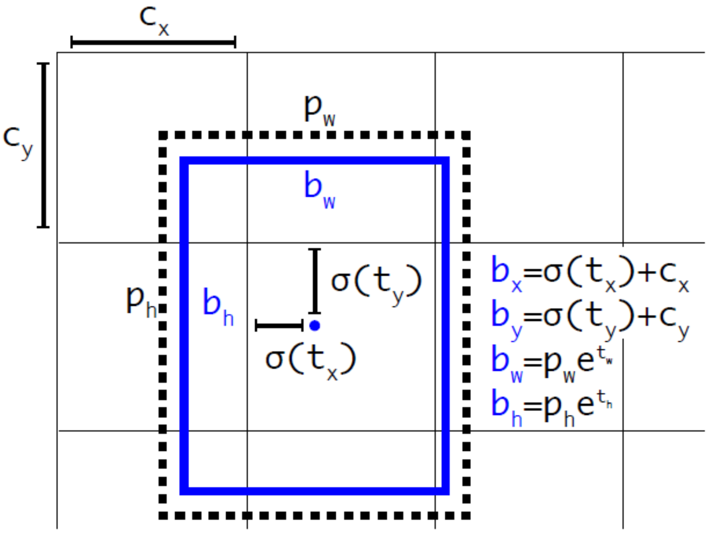


Figure 24 Bounding box with the priors pw and ph along with the predicted width and height [58].

In this model, the output layer uses a linear model while all the other layer uses ReLU activation. The YOLOv3 network predicts 4 coordinates, bx, by, bw, bh for each of the bounding boxes, and the predictions are shown in Figure 24.

3.2 Soccer field detection (Field modeling)

Field modeling is an essential step for detecting and tracking ball and players. It is also helpful in building systems such as Hawkeye detection. Methods that can be useful for field modeling are described below.

3.2.1 RGB to Grey

The first step is to convert the image from RGB to Greyscale. This can be done by, a weighted combination of RGB channels(equation 2).

$$Z= 0.299R + 0.587G + 0.114B \qquad (2)$$

where, R G & B is the color planes of Red, Green, and Blue colors and Z is the resultant grayscale value.

This conversion is necessary as, the RGB image consists of three separate(i.e Red, green, and Blue) 8-bit channels while on the other hand Greyscale image consists of a single channel of 8-bit where the pixel value of each pixel range from 0 to 255 [81]. Converting the image from RGB to greyscale helps in reducing the complexity.

3.2.2 Canny edge detection

Canny edge detection is a widely used edge detection algorithm that utilizes a multi-stage algorithm to detect various kinds of edges in the images [82]. As this algorithm is based on the grey scale so the first step for using this algorithm is to convert the image to a greyscale image, it further consists of the following four stages [83]:

I-Noise reduction

II-Gradient calculation

III-Non-maximum suppression

IV-Double threshold

I- **Noise reduction**- Noise reduction is done by using Gaussian blur (explained in section 3.1.b).

II- **Gradient calculation**- this step is used to detect the edge intensity and direction by calculating the gradient of the image. A change in pixel intensity represents the edge and to calculate this change in intensity some filters such as the Sobel edge detection filter can be used in both horizontal(x) and vertical(y).

By using the Sobel filter at input a clear and noise-free image at the output can be obtained. Sobel filter scans the image in the x and y direction with a standard 3X3 kernel (Figure 25). As the image got scanned in a sliding window manner by using this standard kernel, any detection of the sharp increase in color intensity is considered an edge[84], [43].

$$\boldsymbol{K_x} = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, \boldsymbol{K_y} = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}.$$

Figure 25 Sobel filters for the horizontal and vertical directions [81]

Magnitude ***G*** and the slope ***θ*** of the gradient are calculated by using equation 3.

$$|G| = \sqrt{I_x^2 + I_y^2}$$
$$\theta\ (x, y) = \arctan\ (\frac{I_x}{I_y}) \quad (3)$$

The result of the Sobel filter on the image can be seen in figure 27 (before applying figure 26). There are a large number of lines being detected with a lot of noise and also a lot of thick lines with the greatest magnitude. Whereas it's the thin lines that are desirable as it gives a better idea of where the edges are and gives a clearer outline of the object. The above mention issues can be dealt with by applying non-maximum suppression as the next step.

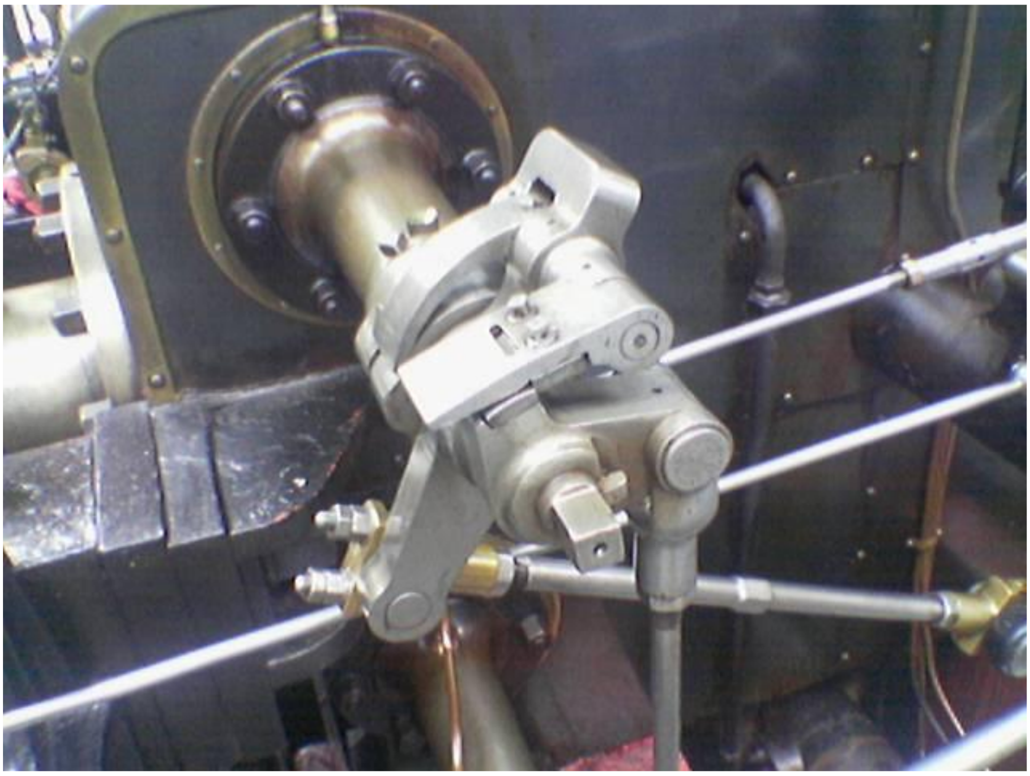


Figure 26 Image before applying Sobel filter [83]

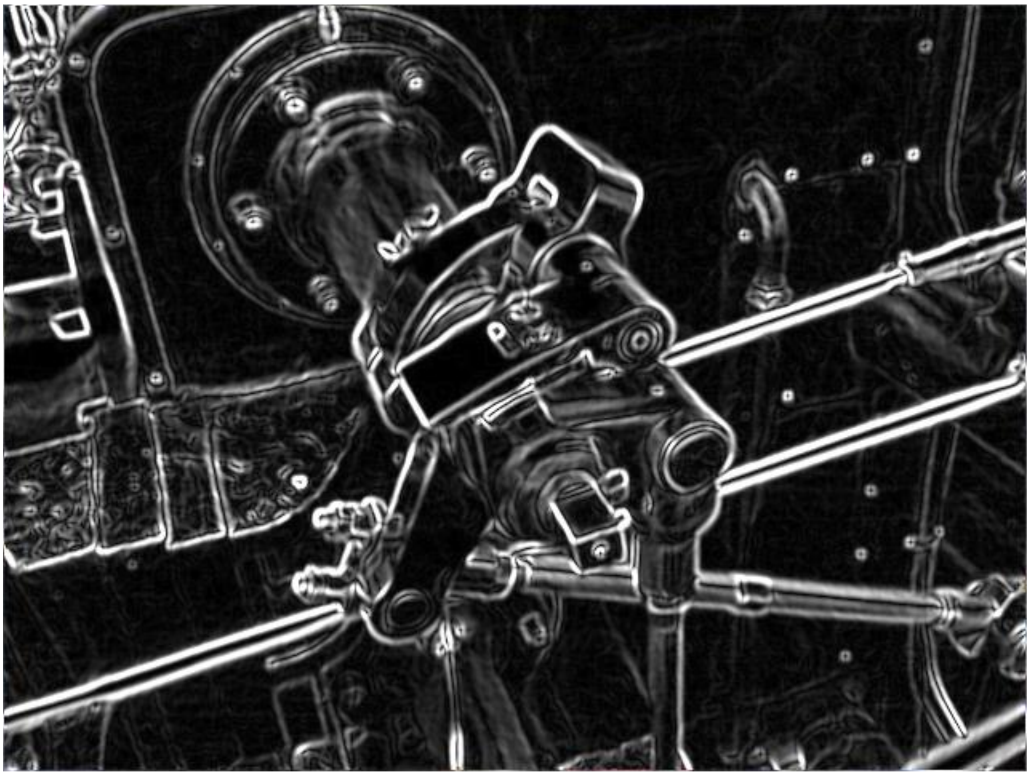


Figure 27 Image after applying Sobel filter [83]

III- **Non-maximum suppression-** Ideally the resultant image should have thin lines therefore to thin out the edges non-maximum suppression method should be used. This method works by only keeping brighter pixels while discarding darker pixels. However, even after applying this step, some variation will remain in the images i.e. some pixels could be brighter than others. So this variation can be removed in the last step.

IV- **Double thresholding**

This method can be used to identify three kinds of pixels as follows

- Strong pixel- these pixels are so high in intensity that it's obvious that these pixels contribute to the final edge.
- Weak pixels- This pixel intensity is not as strong as the strong pixel intensity so it is not granted that if they contribute to the final line, but then they are also not weak enough to get discarded.
- Others- These pixels are considered non-relevant for the image

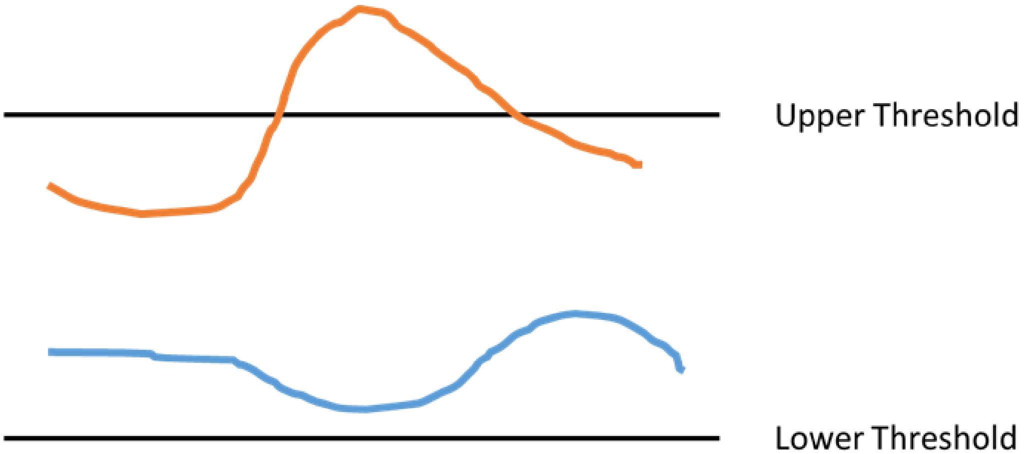


Figure 28 The blue line will be filtered out while the Orange line will be kept.

As the name, double threshold suggested, the two magnitudes are defined as shown in figure 28. The upper threshold is for the pixel that is to be considered as the edge, the lower threshold is for the pixel which is below the line and will be discarded. And if the pixels fall between these thresholds, then the pixel will be kept only if it is connected to the other pixels of the above threshold, otherwise, it will get filtered out. Figure 29 shows the resulting image after applying all the steps of canny edge detection. From the final image, it is clear that although a lot of noise has been filtered out but still the edges that are needed to define the image are still intact.

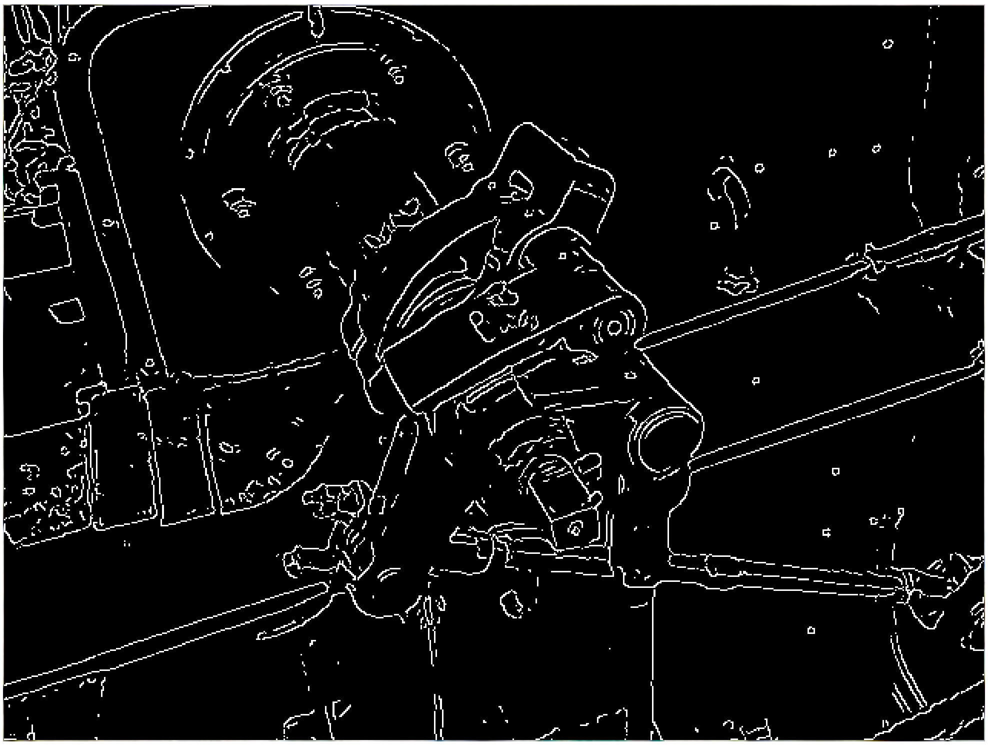


Figure 29 Final image after all Canny edge detection steps

3.2.3 Hough line detection

Hough transform is one of the convenient ways of detecting objects that move in straight lines in the video [85]. This method is mostly used to convert a set of collinear points into a form that helps the threshold operator to detect the moving object and calculate the angle and radius of its trajectory concerning the axis of origin of the image(frames of the video) [34].

Mathematically a straight line can be represented by equation 4.

$$Y = mX + b \qquad (4)$$

Where m is the slope of the line whereas b is the point of intercept of the line at the y-axis. While in the case of a vertical line slope of the line i.e. “m” will be unbounded. To manage such situations Hesse's normal form of representation of line is used(equation 5).

$$r = x\cos(\theta) + \ y\sin(\theta) \qquad (5)$$

Usually, before applying the Hough transformation, edges are detected through some edge detection method after which only the edge of the objects in the image is left. Then every pixel of the remaining edges is scanned. A pixel is a single point in the image space where is the possibility of the passing of an infinite number of lines. As a line is represented as a single point in the Hough space. So if the infinite number of lines passes through a point it will be represented as sinusoidal in Hough space (Figure 30). While analyzing all the pixels, the Hough space will be filled with many overlapping sinusoidal. And each of these intersections can be considered the candidate for a line. As the intersection point at the Hough, space is represented as a straight line in image space, many intersections in the Hough space represent that the image space consists of multiple pixels that have the same line passing through them.

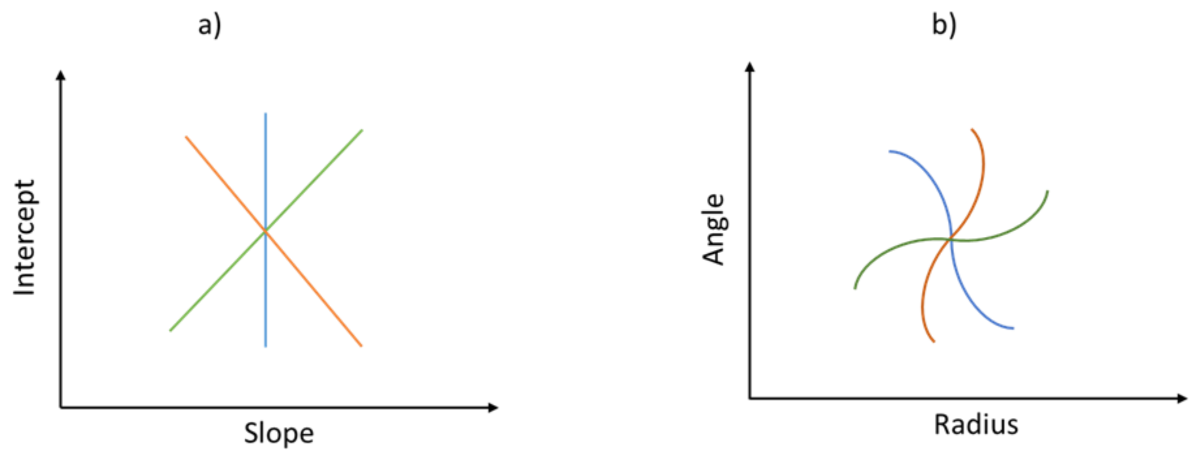


Figure 30 Hough transform of collinear points. (a) slope-intercept form (b) Angle radius form

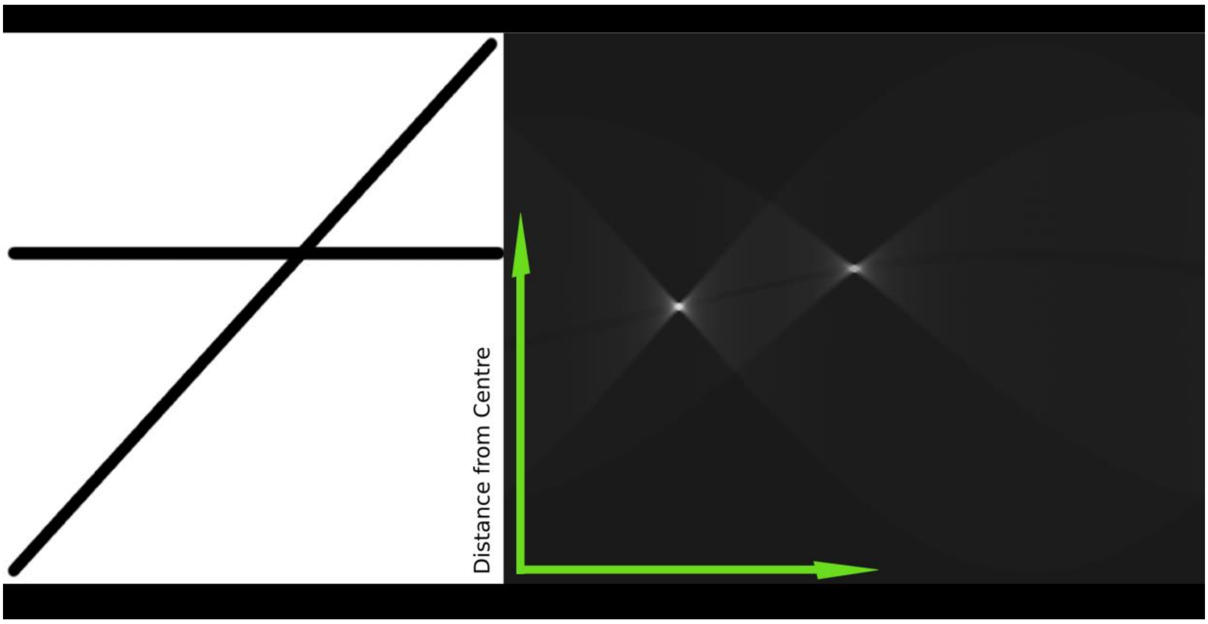


Figure 31 Hough space representation of two lines in the image space[83]

Figure 31 represents the Hough space in which two bright points are seen, these bright points are two lines in the image space. More the intersection of the sinusoidal the brighter spot it becomes. By using the library such as OpenCV, a minimum threshold can be applied to choose the line with the higher number of sinusoid intersections. After scanning all the pixels in the image space, it's become feasible to extract the line by considering only the brightest spots in the Hough space. When instead of using all the edge points, only randomly picked edge points are considered from the image then this form of Hough transformation is called Probabilistic Hough transformation. By using this transform accuracy of the model remain high with a significant reduction in execution time.

Based on the understanding of the aforementioned possible methods for the ball and player detection, three models are selected based on their suitability for the eSport football match. These three methods are described in chapter 5 and the most suitable method is selected for the analysis. Before analyzing the game, it is of utmost importance that the data processing has to be accurate then only a meaningful analysis can be performed. The data processing related details are provided in the next chapter.

3.3 Summary

The eSports games especially FIFA are gaining popularity among people and there is a need for computer vision based analysis to improve understanding of player tactics and strategies. Due to similarities between esports Football and regular football game, several computer vision based methods used in regular football game analysis can be utilized for the analysis of an eFootball game. For an accurate analysis. the tracking of fast-moving objects such as a ball and players is of utmost importance. For ball detection and tracking, there are several challenges such as fast and random movements, occlusion by the players or field features, change of ball shape due to camera movements, etc. For player detection and tracking, occlusion by the other players is a major challenge. Another important feature is the mapping of the play-field which is another major challenge as the accuracy of this feature determines the accuracy of the analysis. In play-field modeling, major challenges come from the occlusion of field lines by the players or sometimes by the ball and also the quality of the video. Due to the occlusion, the field line can be detected as 2 or more separate lines which can reduce the accuracy of the analysis. For a successful computer vision-based analysis, several studies related to different games such as football, handball, racketball, cricket, tennis, etc. have been performed and these studies provided useful insight into the possible challenges and mitigation strategies. These results are studied and their finding is used in understanding the possible methods for the analysis.

3.4 Conclusions of the current research

According to the literature study, three methods are found to be more suitable for the detection and tracking of the ball and the players in an esports Football video analysis. These methods are 1) color-based detection and tracking, 2) template-based detection and tracking, and 3) deep learning (Yolov3 model) based detection and tracking methods. All three are suitable based on the video and game conditions. The color-based detection technique is quite a significant method in the field of object detection. It is very simple as well as an efficient way to represent and match images based on the color histogram. This method has a limitation in the situations when illumination circumstances are not equal, as in this condition accuracy degrades significantly. Template-based matching is another important technique in the computer vision field in which a template of the target object is used to detect and track the object in the image. The template is matched in the image and where there is the maximum probability of the matched object. But this method also has some limitations such as, it is more efficient when the images are of larger resolution and also sliding the template image in a very large image with one pixel at a time could be computationally very expensive as well. The third chosen method i.e. YOLO (You Only Look Once) object detection is a deep learning based object detection method. This method is quite efficient, especially with the availability of pretrained coco dataset based model which makes the object detection model building process easy and quick. The object detection on the customized dataset is quite challenging as data annotation could be a tedious process depending upon the requirement and also training the data, especially with multiple classes can take from days to weeks.

In the next chapter, the data processing required for the analysis is explained.

4. Data Processing

In this project, only one downloaded video from gameplay by an eSport game Youtuber of FIFA 2022 is used for analysis. It includes a single camera that moves horizontally and vertically. The video has a bigger field of the actual gameplay and at the bottom of the screen, the mini-map of the game shows the entire field at once. Figure 32 shows the bigger field on the left and the zoomed-out mini-map on the right side.

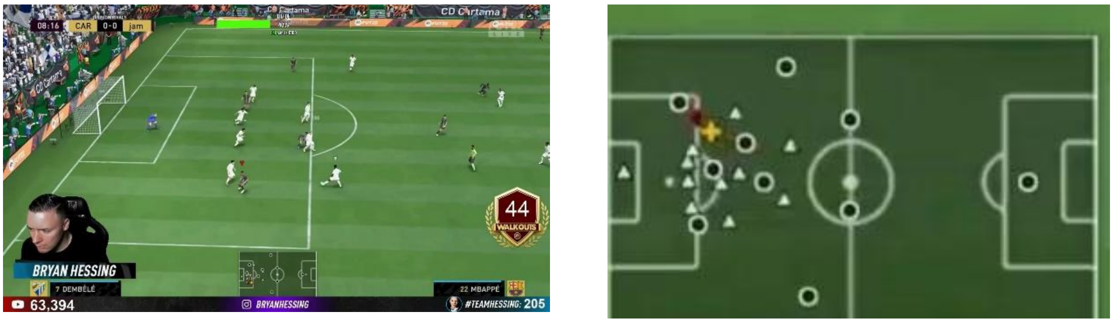


Figure 32 Bigger field (left) and mini-map (right)

The data is read frame by frame from the video in the form of images. Due to the camera movement in the horizontal and vertical directions, there is no single origin of these images of the bigger field while the location of the mini-map on the screen(image) almost remains the same.

4.1 Data Preparation

In this research, three different models have been developed therefore data preparation is also performed separately for these three models. The data preparation for these models is described below.

4.1.1 For color-based detection and tracking

The video is read frame by frame (in the form of images) using the OpenCV library. Afterward, these images are provided as the input to the model. For the detection and tracking on the bigger field, the entire image is considered while for the mini-map, the region of interest i.e. the mini-map area on the image is zoomed out on each frame. These images of mini-maps are given as the input.

Figure 33 Templates used in the template matching model

4.1.2 For template matching based detection

For template matching, two kinds of data (source image and template image) are required as input. For the source image, similar steps as in the case of color-based detection on bigger field images are performed and the entire image is given as the source image. While for the template image, some of the images of the players and ball are given as input. The trial and error-based template matching which is giving the best results are taken as template input. Figure 33 shows the template used for player and ball detection.

4.1.3 YOLO based object detection

In YOLO based detection process a large number of training data points are required. For this purpose, around 2200 images are used to train this model. These images are taken from 7 different videos having players with different types of jersey colors.

4.1.3.1 Data annotation

The data annotation is performed by using the LabelImg which is a graphical image annotation tool it labels object bounding boxes in images. Based on most of the available choices of jersey colors they are categorized into 5 main colors i.e. Green, blue, red, grey, and white. Different shades of the same color are included in the main color category such as orange and pink, and red and white are included in the same red color category.

Figure 34 shows two different categories of blue categorized into single blue while figure 35 shows different kinds of red that were categorized into red categories

Figure 34 Different shades of blue categorized in blue color

Figure 35 Different shades of red are categorized in the red category.

5. Methodology

This chapter describes the methodology used to answer the research questions posed in the introduction chapter. The necessary background information pertinent to the possible methods to be used is already described in chapter 3.

5.1 Ball and player detection and tracking

This section describes the methodology used for the ball and player detection and field modeling. It focuses on the three selected methods: 1) color-based, 2) template matching and 3) YOLO-based tracking, from the literature for the ball and player detection and tracking. These methods are adequate for answering the first research question “How to accurately detect and track ball and players in an eSport FIFA 2022 game video?”. Based on the literature review, the following models are proposed to extract the ball and player detection and tracking and play-field modeling

5.1.1 Color-based ball and player detection and tracking

In this method, the detection of the ball and the players is carried out based on their color. Figure 36 shows the steps of the complete model. In the case of occlusion, Kalman’s filter can be used for continuous detection and tracking. This method has been implemented on both larger field as well as mini-map. First Kalman’s filter is initialized along with its matrices, then the video is read on a frame-by-frame basis. On each of these frames following steps are implemented.

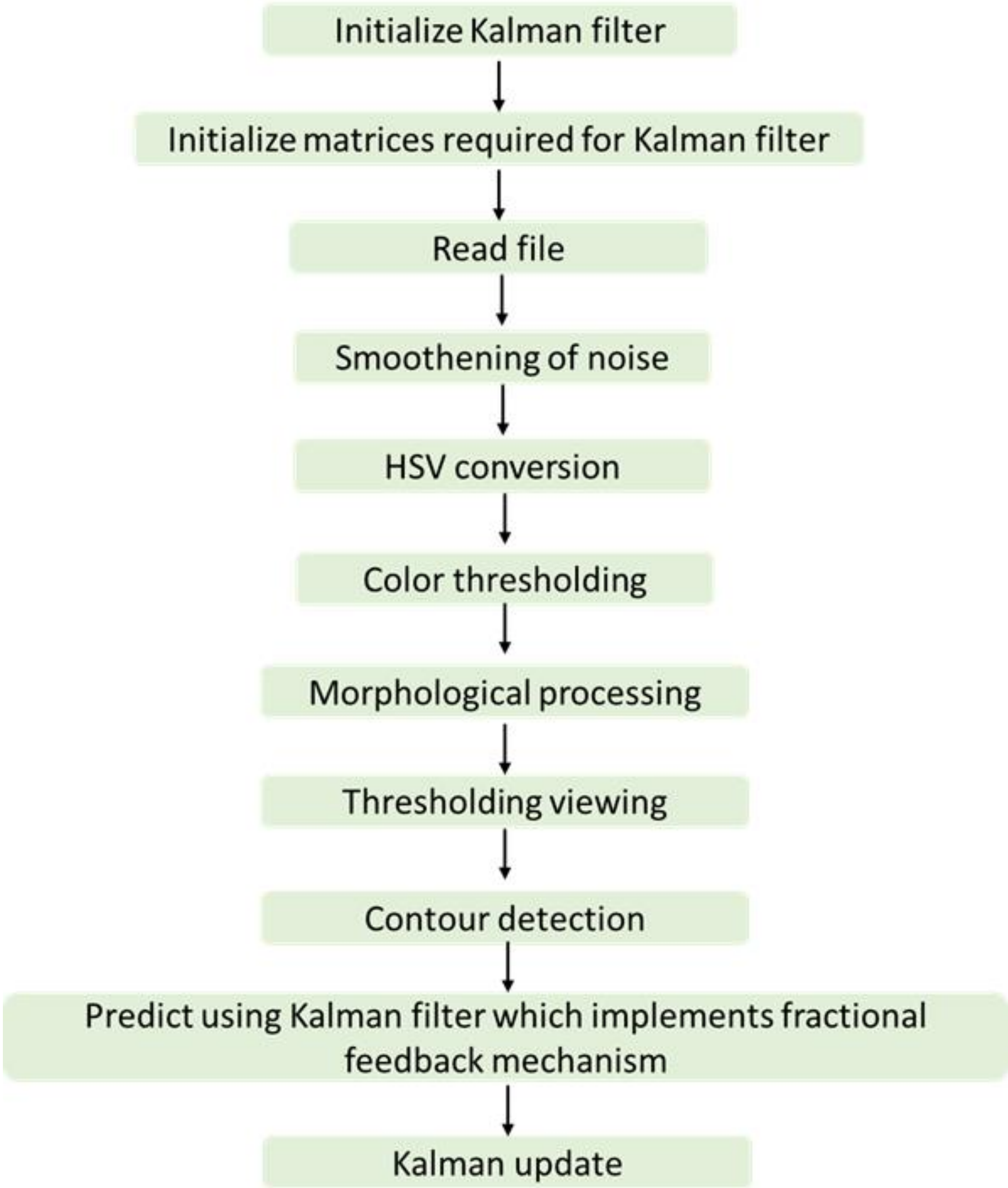
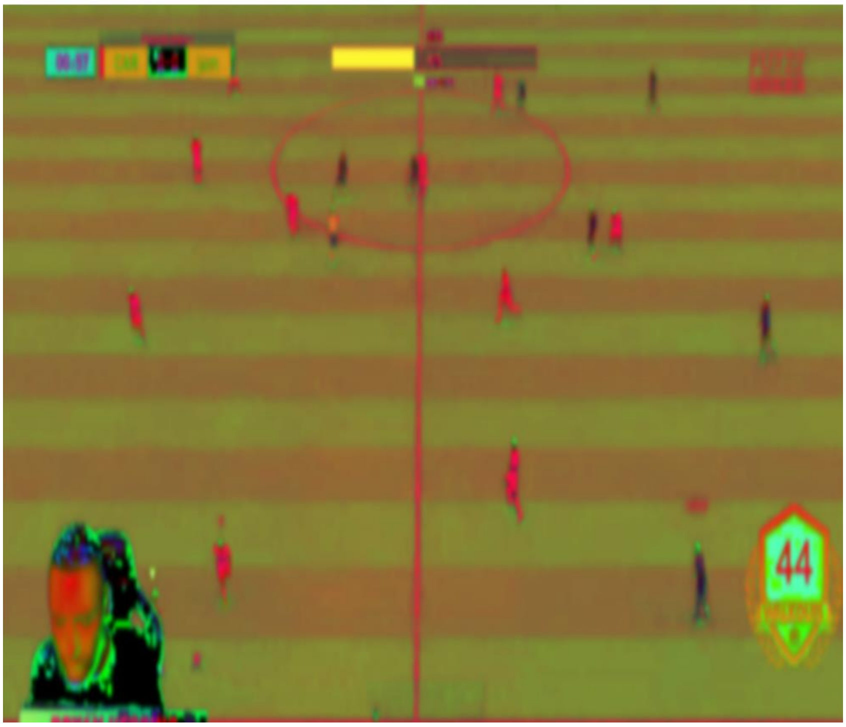


Figure 36 Color based ball and player detection flow chart

5.1.1.1 Gaussian Blur And Color Filtering

To eliminate the noise from the frame, the noise smoothening can be done by using the function of Gaussian blur. Figure 37 a) and b) show the effect of gaussian blur on the original image. Next, this blur image should be converted from RGB space to HSV space as Hue Saturation Value is used to separate image luminance from color information and it is useful in the situation where the color description is important. Then several color thresholds can be initiated stating their lower bound and upper bound values to detect the color of the ball and soccer team according to their jersey color.

a)



b)

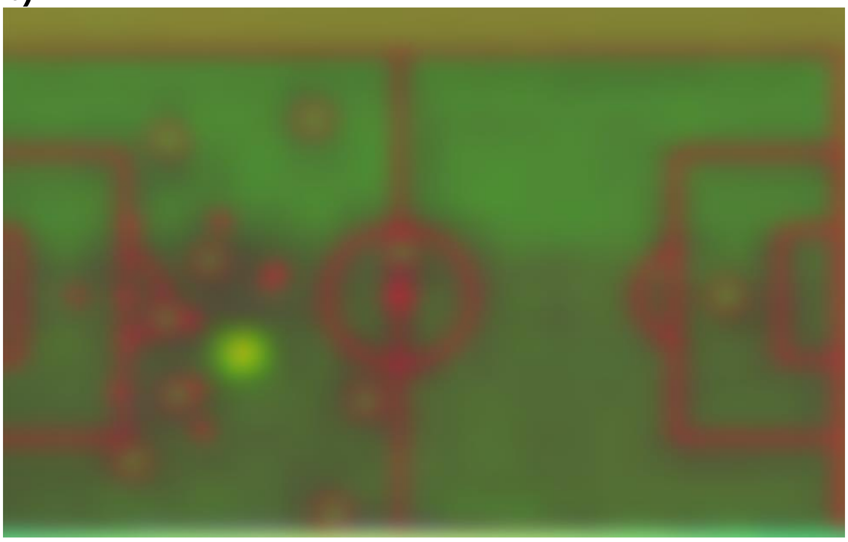


Figure 37 Image after Gaussian blur in a) bigger field b) mini-map

5.1.1.2 Morphological Operations

Now the resulting image is converted into a binary image and morphological operations such as erosion and dilation can be performed on this binary image. These morphological operations are useful in closing any loose end in the image or filling the gap that has occurred due to the noise in the image. Figure 38 a) and b) shows morphological operations on the resulting image for bigger field and mini-map respectively.

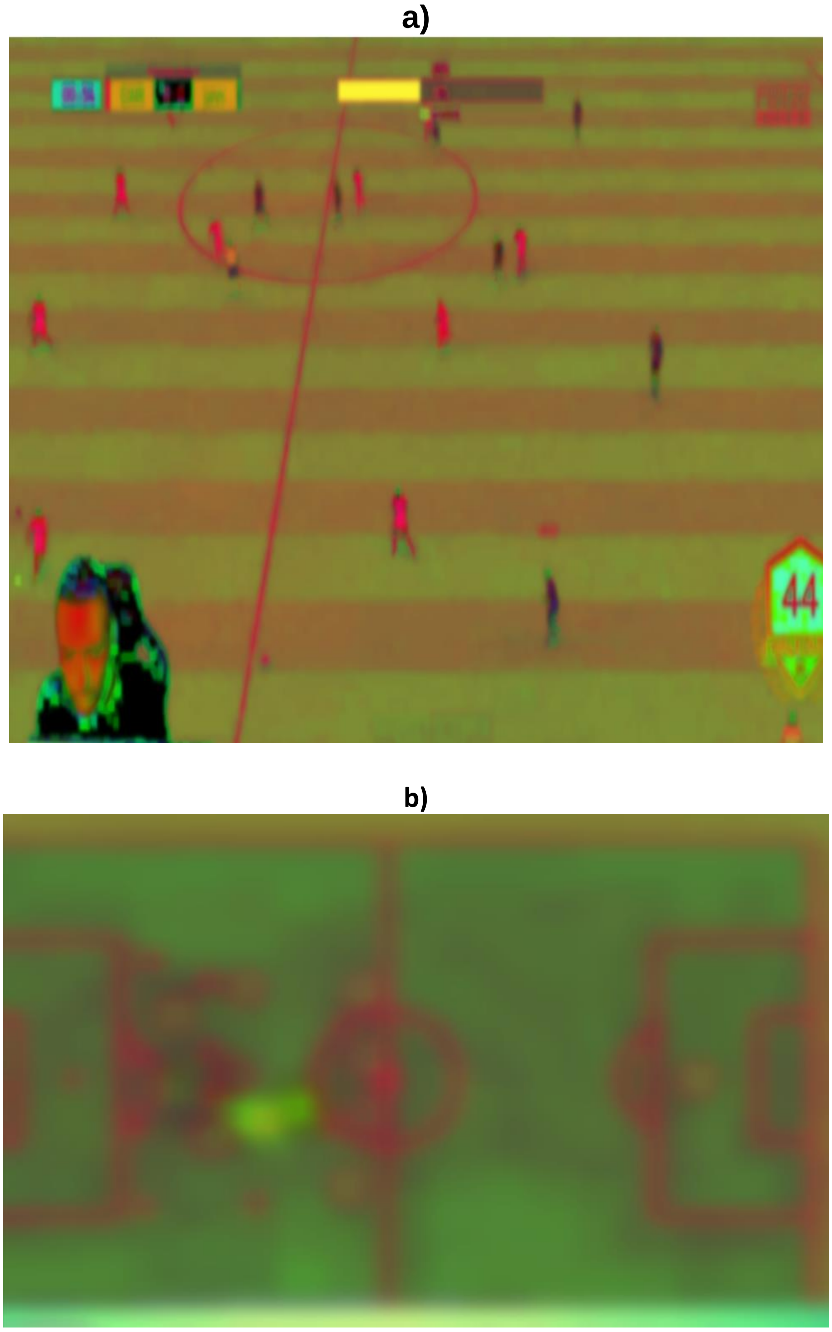


Figure 38 Image after performing morphological operation a) bigger field b) mini-map

5.1.1.3 Contour Detection

After the morphological operation, the contour detection is the final stage of the ball and player detection where the contour size of the ball should be small as compared to the soccer player contour size.

5.1.1.4 Kalman's Filter

After detecting the ball and players, Kalman's filter is used for the prediction of their location in case of occlusions. The detection using this filter is shown with a bounding box around the tracked object. Figure 39 shows the detected ball using Kalman's filter using a blue circle box.

Figure 39 Ball detection with Kalman's filter

5.1.1.5 The final outcome of the color-based model

This model has applied to both the bigger field as well as to the mini-map. Unfortunately, it is not able to accurately detect the ball in the bigger field due to changes in the ball contours because of the changing distance between the ball and the camera. It provided mostly false positive outcomes from this detection for the ball (figure 40 a). Though it provided accurate detection of players due to limited changes in the player contours. One example is shown in figure 40 b

a)

b)

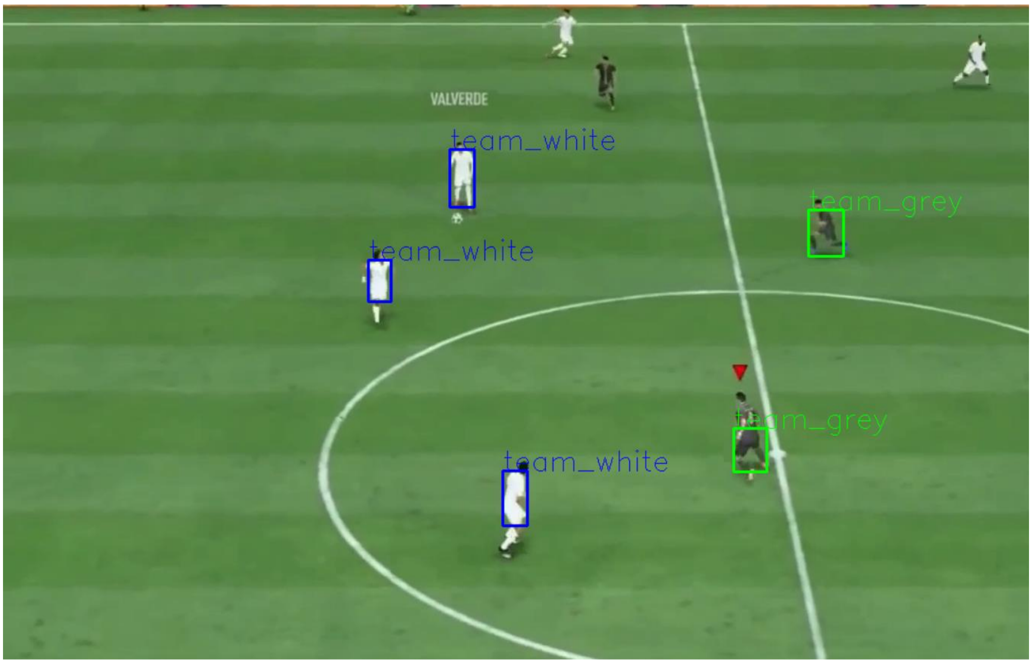


Figure 40 a) False positives for the ball detection and accurate player detection in the bigger field using this method b) Player detection using the present method.

5.1.2 Template-based ball and player detection and tracking

Figure 41 shows the complete architecture of the template matching algorithm in the source image. The starting few steps of this model are similar to the previous model in which the Kalman filter and its matrices should be initialized in the beginning. Then the source video can be read per-frame basis, these frames can be converted from RGB to greyscale (figure 42 a) and the converted frames can be smoothened for noise by using a Gaussian filter (figure 42 b). This conversion is done because computation is easier in the greyscale as compared to RGB color space. After all, RGB consists of 3 channels i.e. Red, Green, and Blue while Greyscale only has one channel. Similarly, the template image should also be added and smoothened for noise and then should be converted into a Greyscale. Now this template image can be searched on the source image. This search will be performed by comparing one pixel at a time of source and template image. On detection of the object in the source image, further prediction can be made using Kalman's filter and bounding boxes can be drawn across the detected object. In this project, the objects to be detected are the ball and the players.

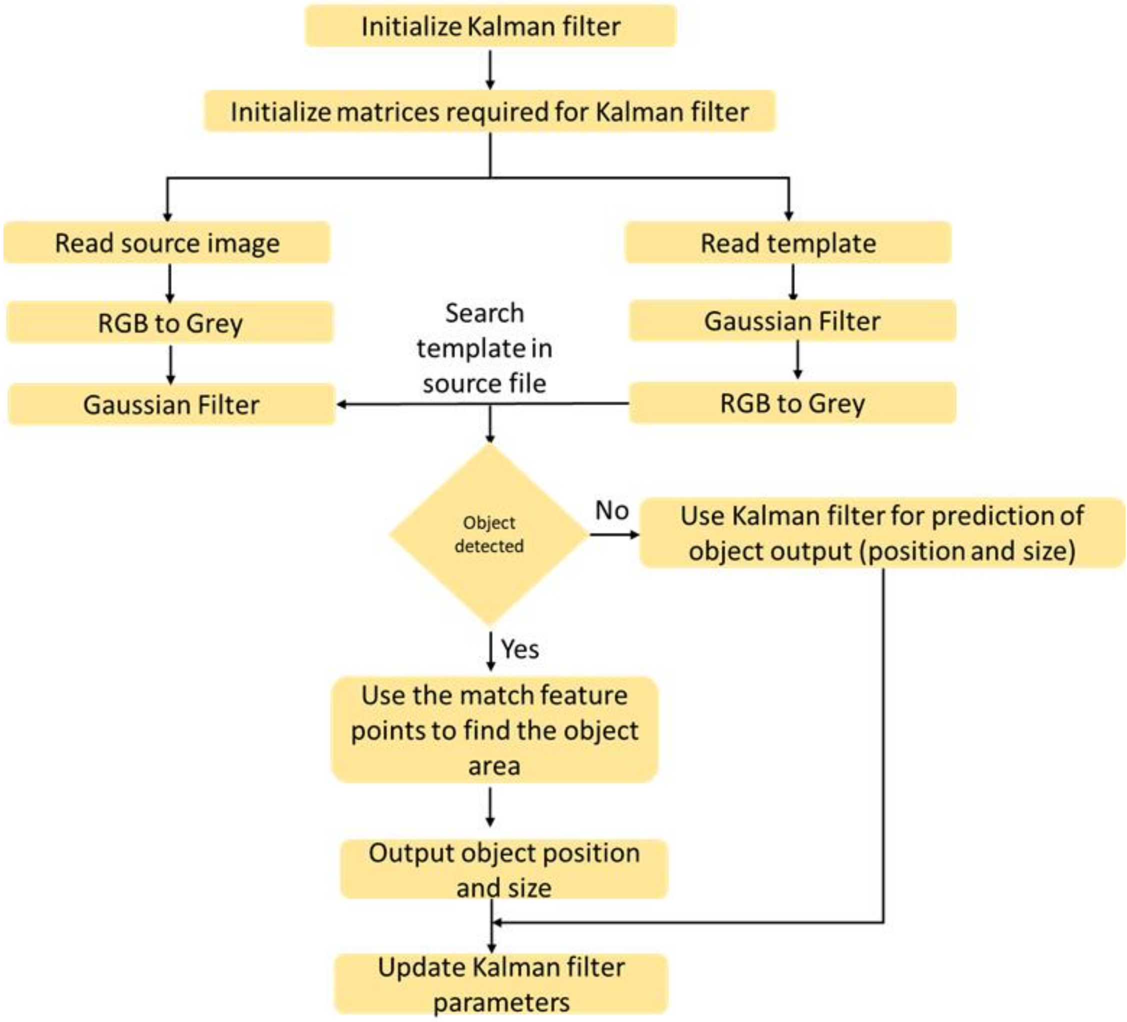


Figure 41 Complete architecture of template matching in the source image

a)

b)

Figure 42 a) Image conversion from RGB to Greyscale, b) Blurred image of the grey scale to remove the noise

5.1.2.1 The final outcome of the template-based model

In this model, the biggest problem is the false detection of the ball. The Player's shoes and an arrow above the head of the player in possession of the ball are falsely detected as the ball and also the ball detection is also not continuous. Figure 43 on the left shows the correct detection of the ball while the figure on the right shows player socks falsely detecting as a ball. Due to huge false positive numbers, this method is not continued for further analysis.

Figure 43 Correct ball detection (left) and incorrect ball detection (right) by the template matching.

5.1.3 Deep-learning based (Yolov3 model) ball and player detection and tracking

In this model, only the last layer of the Yolo can be trained by using the annotated data containing the ball and team color information. Now a threshold on the confidence of detection of the object can be set. For example, if the confidence threshold is set to be 20 % then objects detected with more than this confidence will consider, otherwise, they will reject. After object detection bounding boxes are drawn across the object and based on the x and y location of these bounding boxes the desired object (in our case ball and player) can be tracked. Figure 44 shows the final output of the YOLO model detecting players and the ball.

5.1.3.1 The final outcome of the YOLO model

The YOLO model is successfully able to detect both player and ball without any false positive detection. This detection is done in the bigger field. But for the tracking of the ball and player, playfield modeling is required which is quite challenging in the case of bigger field due to camera movements during the game.

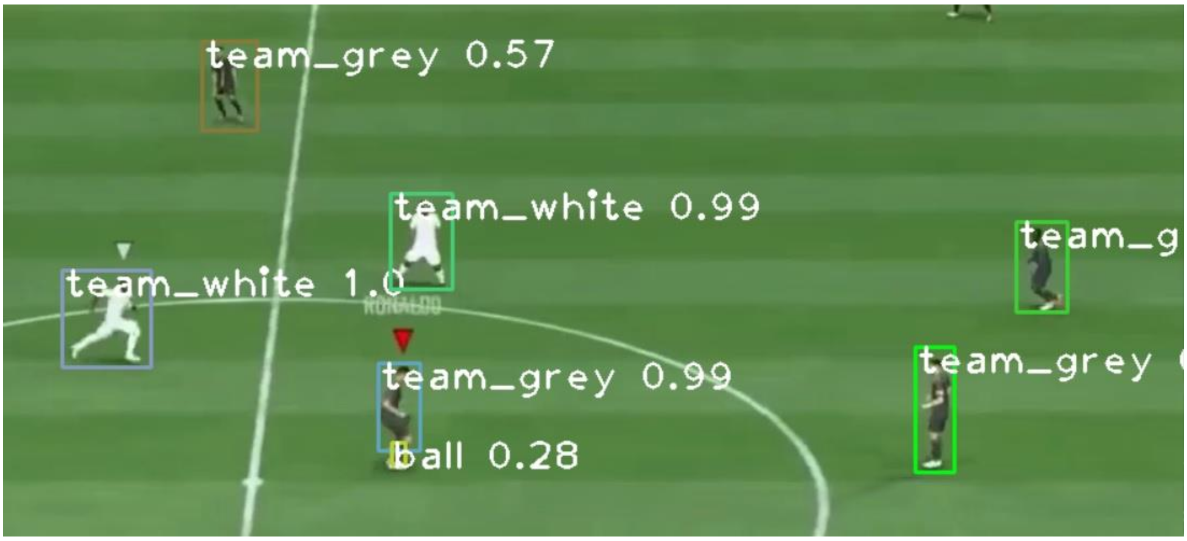


Figure 44 The ball and player detection by the YOLO model

5.1.4 Combined Model

The final model consists of a combination of model 1 color-based ball detection from mini-map and model 3 Yolo-based player and ball detection. Both models are good at detecting ball and the Yolo model is also additionally capable of detecting players as well. To track the ball and plotting data mini-map is better as no additional field modeling is required. Although the YOLO model is good at detecting and tracking the ball and the players, plotting the data from this model is challenging due to the movement of the camera which changes the origin of each frame of the bigger field.

Below are the steps of the combined model

1) Detect the ball and the player at the bigger field frame-wise.

2) Calculate the player in the possession of the ball by estimating the minimum distance between the ball and the player in the respective frame(the player with minimum distance to the ball is considered as the player in possession)

3) Now at that frame look for the ball position on the mini-map.

4) Final output - providing ball and player location and player in possession information.

Although the combined model performs well in detecting and tracking the ball, still in the case when the player is in possession of the ball is occluded by the other player then the model is not capable of tracking at that instance. In this study, the occlusion of the player has not cause much effect on the outcome. Even after the occlusion and temporary non-tracking of the ball and the player in possession, enough data points are generated to track the trajectory of the ball and the players.

5.1.5 Model Validation

The combined model is validated using the data from the video. Figures 45 a) and b) show the location of the ball and the player in possession in the actual video (left) and the mini-map (right). The location in the mini-map is determined by using the combined model. In both cases, the ball and player location are detected correctly by using the combined model. The successful validation of the combined model answers the first research question related to the accurate detection and tracking of the ball and players in an eSports FIFA 2022 game.

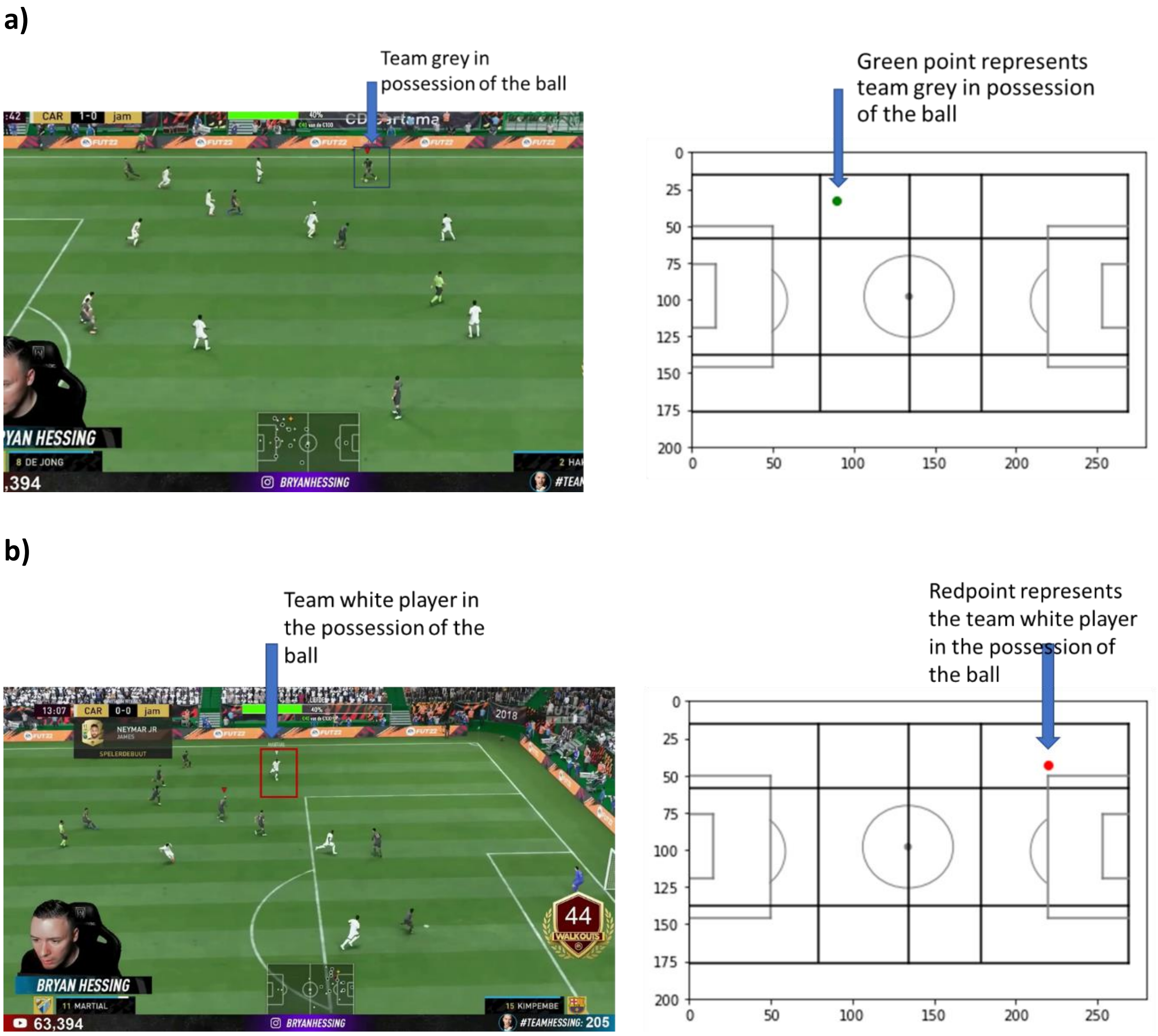


Figure 45 Combined model validation for the grey team (a) and the white team (b)

5.2 eSports player game tactics determination

By using the combined model described in section 4.1.4, the coordinates of the bounding boxes surrounding the ball and the players can be determined. These coordinates can be used to find the centroid coordinates of the players and the ball which is required for further analysis of the game to determine the eSports player game tactics and strategies. The rest of the section describes the proposed process.

5.2.1 Ball possession by the player

The player in the possession of the ball can be determined by calculating the minimum distance of the ball centroid with all the players bounding box centroid detected in that frame and the player within the closest proximity to the ball will be considered as the player in possession. Figure 46 depicts the player in the possession of the ball

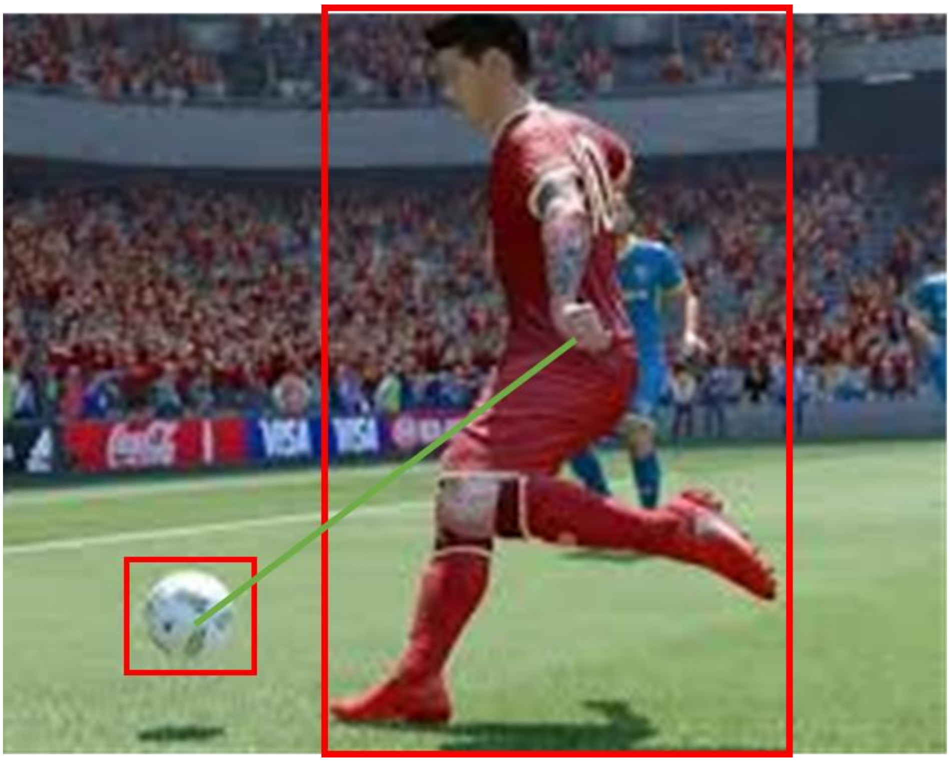


Figure 46 Football player in the possession of the ball

5.2.2 Action Zones

Typically a football field is divided into three action zones, defensive third, mid-field, and attacking third. For this purpose, the field to be analyzed is equally divided using three vertical lines as shown in figure 47. Actions zones can be a good indicator of which eSport player is more dominating during the game. By knowing the ball position in the defensive/attacking thirds, it is possible to find out which team was in the attacking/defensive mode during the game or part of the game.

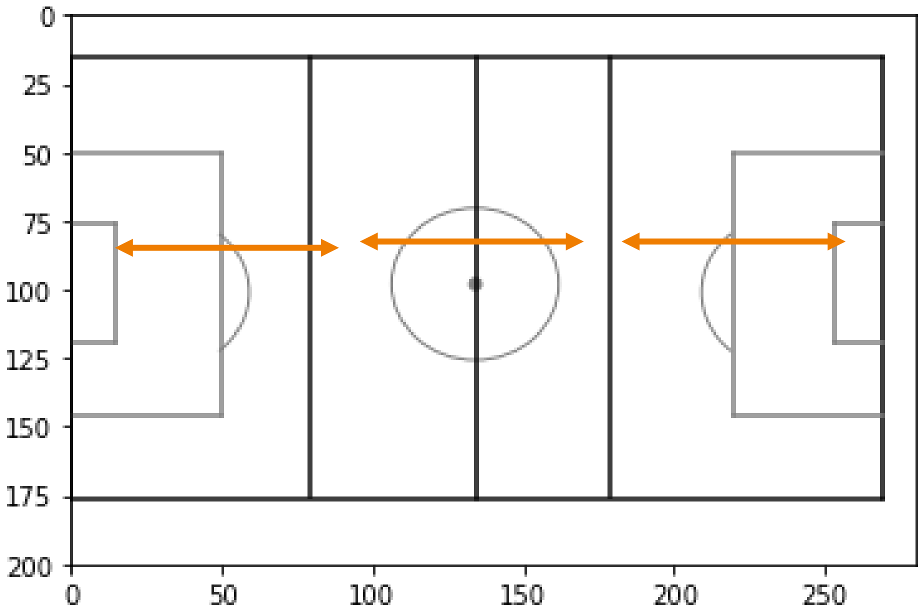


Figure 47 Action zones in the analyzed football field

5.2.3 Attack Zones

Attack zones can be a good indicator of which eSport player is more dominant over others. By knowing the ball position in the defensive/offensive thirds, it is more likely to find out which team was more attacking over the other. They can be a good indicator of knowing which part of the field is mostly used by the attacking team to try to score the goal. These attack zones are created by first dividing the field into two half i.e. left and right halve. And then again divide each half into three equal parts i.e. upper, middle, and lower. The team should be considered attacking when it is in the possession of the ball and also in the opponent’s half. Possession of the ball in their half will not be considered an attack. Figure 48 (a) and (b) show the attack zones of the right and left teams respectively.

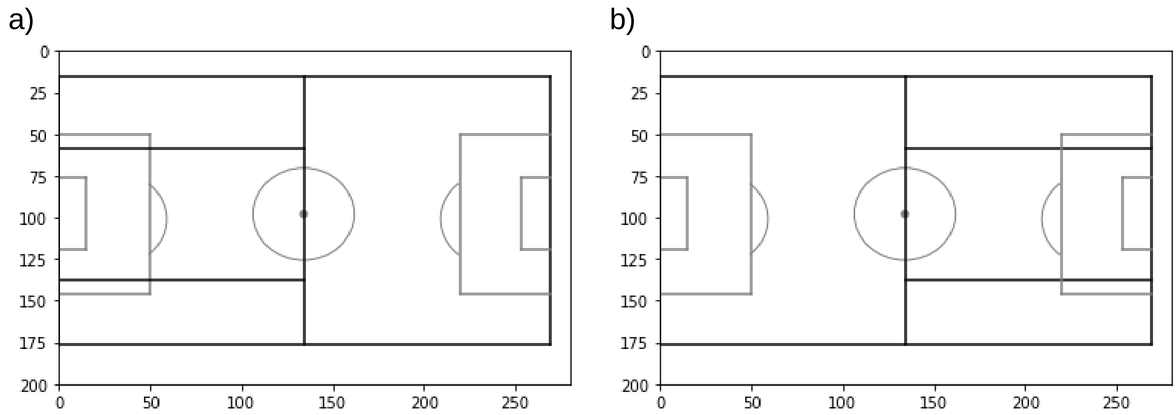


Figure 48 Attack zones for the right team (a) and the left team (b)

5.2.4 Ball position by halves

Another useful statistic is to find the fraction of time in which ball location is in one of the half (left or right) of the field. For this purpose, the field is divided into 2 halves as shown in figure 49 and the ball location is determined in one of the halves. Since the mini-map is used for the determination of these statistics, it is always visible in the field. The location percentages are calculated by counting the number of frames of the ball in each of the half divided by the total number of frames in which the ball is detected.

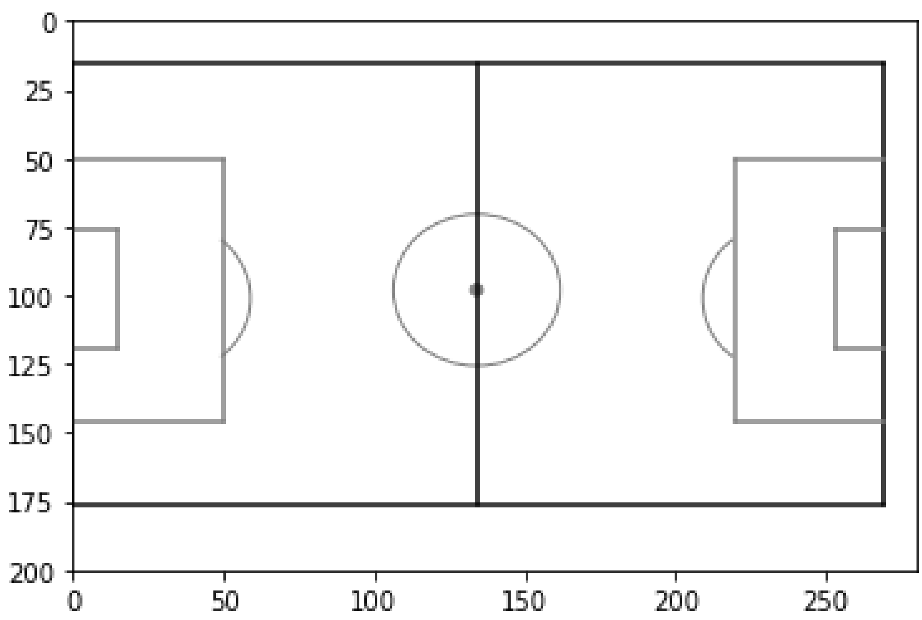


Figure 49 Field division in halves

5.2.5 Goal attempts by each team

To determine the goal attempt, the field can be divided into 12 zones by combining the action and attack zones. The movement of the ball through these zones can be traced in the analysis and it can provide more details related to the goal attempt tactics of the eSports player. Figure 50 shows the division of the football field into 12 zones by combining attack and action zones.

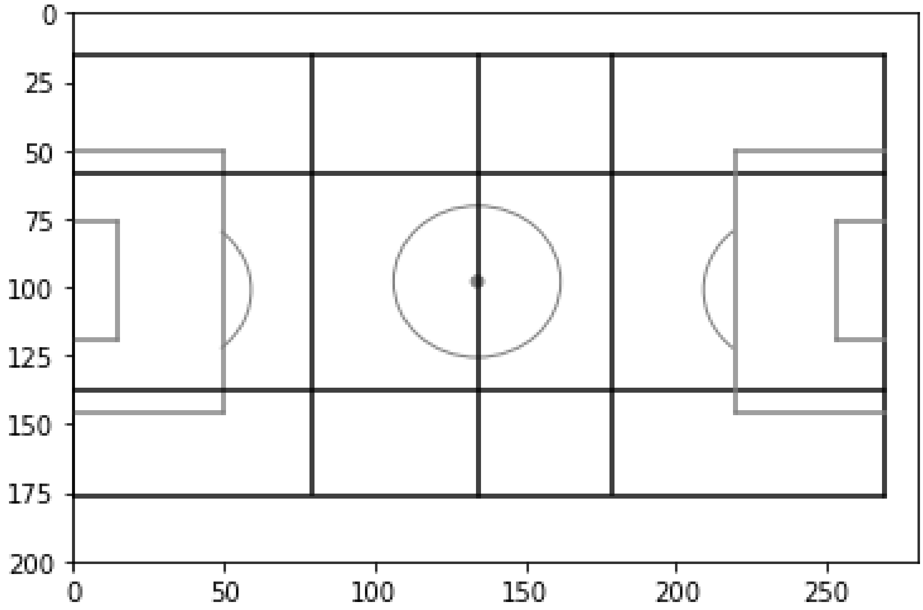


Figure 50 The division of the football field into 12 zones for the goal attempt analysis.

6. Results

This chapter describes the results pertinent to the eSports player game tactics and strategy determination. The methodology described in the previous chapter is used to find the tactics and strategies of the player. These results provided the answer to the second research question “*How the extracted data can be used to determine the game tactics of the player?*”. To find the game tactics of the player, several statistics such as ball possession by the players or teams, ball location in half field, ball locations by the teams in action and attacking zones, and goal attempts are computed. The rest of the section describes these results. For the analysis purpose, only one game video is used in which teams were wearing white and grey uniforms. For clarity purposes, white players are depicted in red color and grey players are shown in green color.

6.1 Action Zone

Figure 51 shows the fraction of time (first half of the game) in which the ball was in different zones. For this calculation, the number of frames in which the ball was detected within a particular zone is counted. It shows that in the first half of the game, most of the time the ball was in the mid-field zone. During the rest of the time, the ball was in zone 1 for relatively more time in comparison to zone 3. It indicates that the green team could be in a slightly more attacking mode as compared to the red team. It is important to note that the attacking mode of a team cannot be judged by just this one statistic and further indicators as explained in the rest of the chapter should be considered before finalizing this mode.

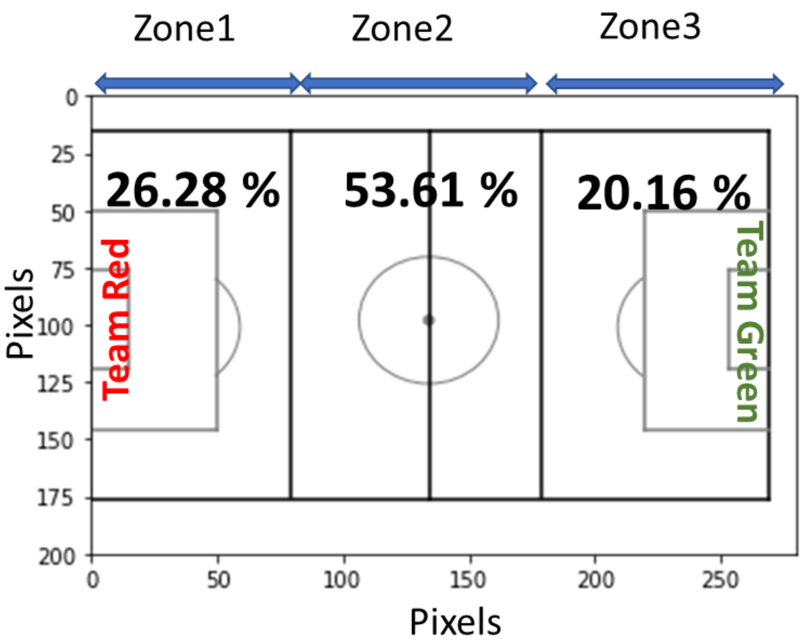


Figure 51 The fraction of time ball is in different zones of the field

6.2 Ball position by halves

In the analyzed video, the ball location is mostly (~ 60 %) in the right half which is related to the green team side (figure 52). This indicates that the red team was relatively more in the attacking mode which contradicts the previous section result (section 6.1). As stated earlier, one result cannot be used to determine the playing strategy of the eSports player. Therefore, other indicators should be considered before finalizing the results.

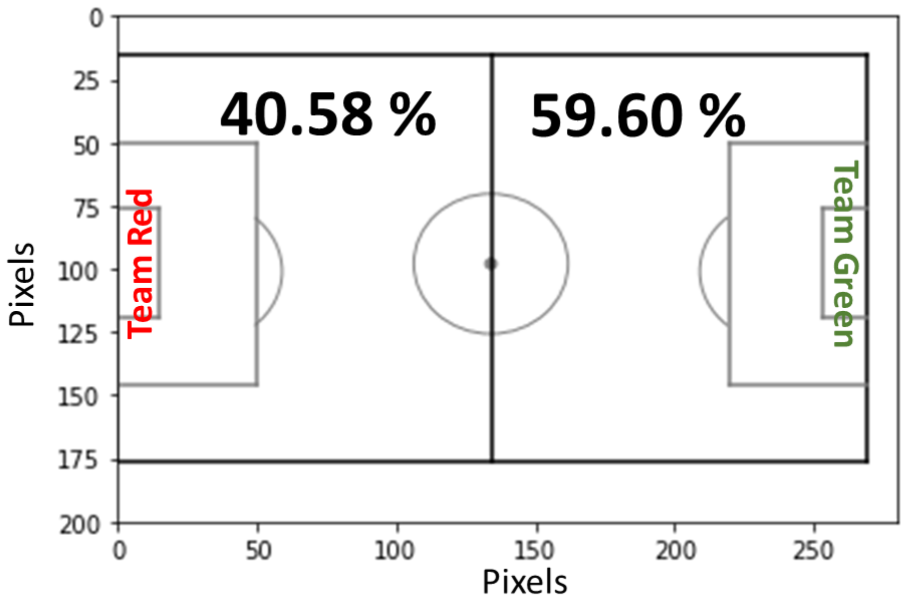
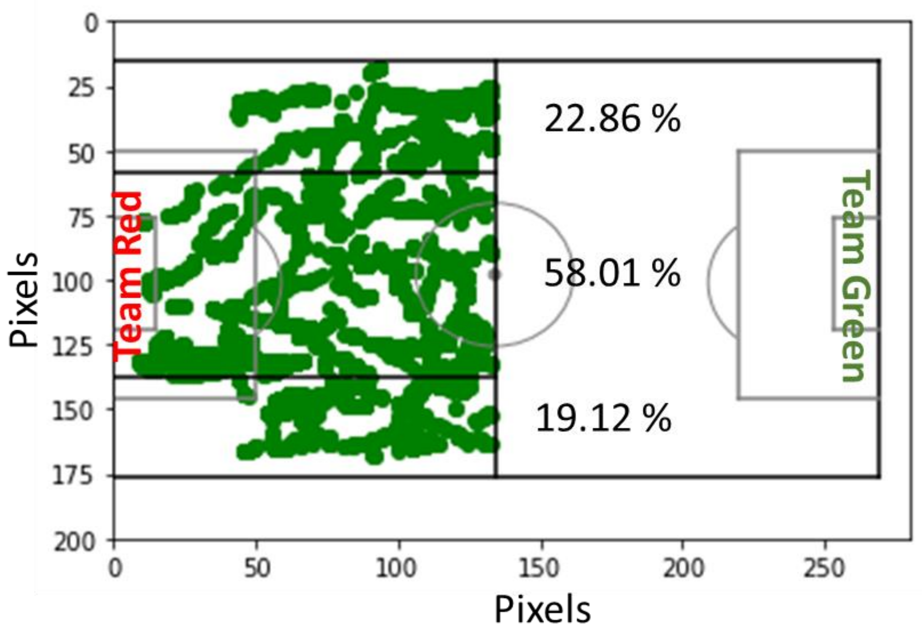


Figure 52 The fraction of ball positions in the halves of the field.

6.3 Attack Zones

As shown in figure 53 a), the green team has mostly used the middle field (~ 58 %) for attacking the red team goal post and their use of the lower field is relatively less as compared to the upper field part. Similarly, the red team has mostly used the upper part of the field (~ 60 %) and their use of the lower part is minimal (only 5.64 %) for attacking purposes (figure 53 b). This result differentiates their attacking strategy in comparison to the green team. By knowing these statistics, the opponent team can try to put more players in the defensive positions in the upper part of the field.

a)



b)

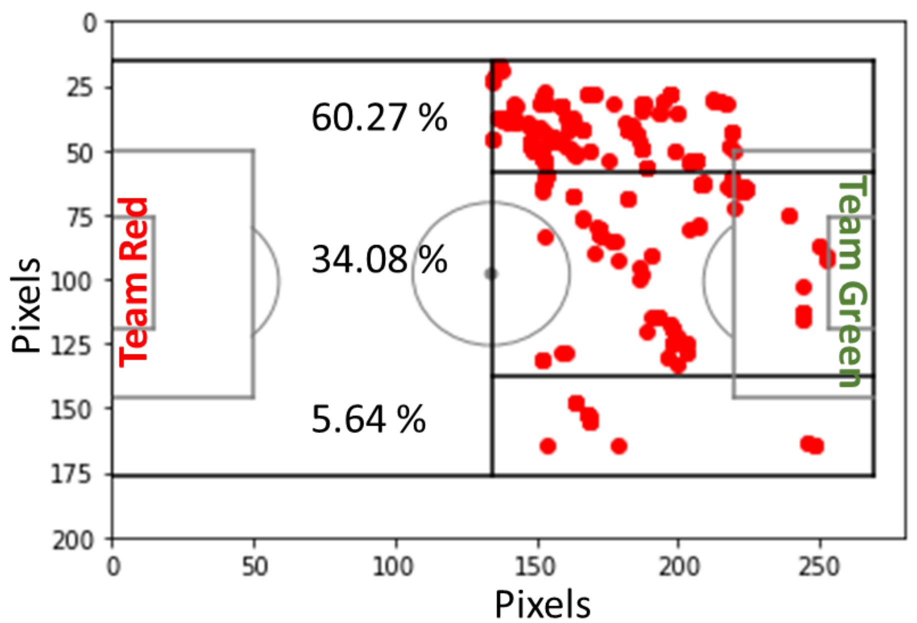


Figure 53 Ball possession statistics in attack zones by a) green team b) red team

6.4 Model validation

For model validation, the ground truth files are generated by taking 50 frames from the test video. Afterward, the intersection of Union (IOU) from the ground truth and prediction bounding box is calculated. The threshold for IOU is set to be 0.5 which means that only the detection objects which have IOU greater than 0.5 are considered true predictions and less than this threshold are considered false predictions. These values are used for the calculation of precision, recall, and f1 score.

Table 2 Precision, Recall, and F1 score of color and YOLO based detection model

	Precision	Recall	F1 score
Color-based model	0.88	0.57	0.69
Yolov3 based model	0.96	0.71	0.80

Table 2 shows the precision, recall, and F1 score of color-based and YOLO-based detection models used in this study. The Yolo-based model has a precision of 0.96 which indicates that out of all detected balls, 96% of the time it was the correct detection which means the model has low false positives. This model has a recall of 0.71 which indicate that out of all the ball in the frames, it can detect it 71% of the time. The precision of the color-based model is .88 which is less than Yolo based model but it is still reasonable. While the recall of the color-based model is .57 which is lower in comparison to the Yolo-based model. The Yolo-based model has a higher F1 score of .80 in comparison to the color-based model which is 0.69, this indicates the performance of the YOLO model is better than the color-based model.

6.5 Goal attempts by each team

For a detailed understanding of the goal done by any team, the ball path for a particular goal is traced for both teams. Figure 54 shows 4 goal attempts by the red team. In these figures, the red line shows the path of the ball before the final goal attempt. Figures 54 a, b, and c are related to the successful goal attempt, and figure 54 d provided the goal path for an unsuccessful goal attempt. The successful goal attempts show that the red team has used all three areas (upper, mid-field, and lower) for proceeding toward the goal.

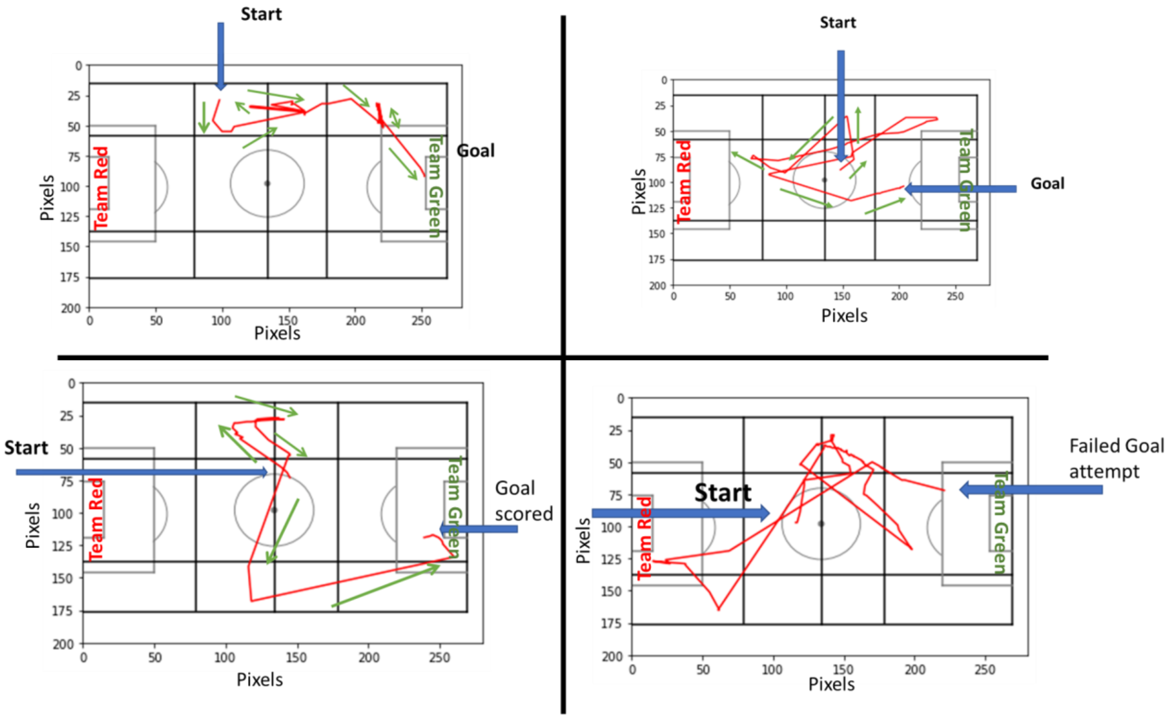


Figure 54 Goal attempts by the red team a) successful, b) successful, c) successful, d) unsuccessful

Figure 55 shows 2 goal attempts by the green team. They only used the lower and middle parts of the field for the goal attempts. Though it might be part of the strategy it cannot be concluded based on just two data points related to the goal attempts. If this model is extended to several games then the goal attempt patterns of each team can be determined up to a certain extent. Although these strategies of the player can also differ from game to game.

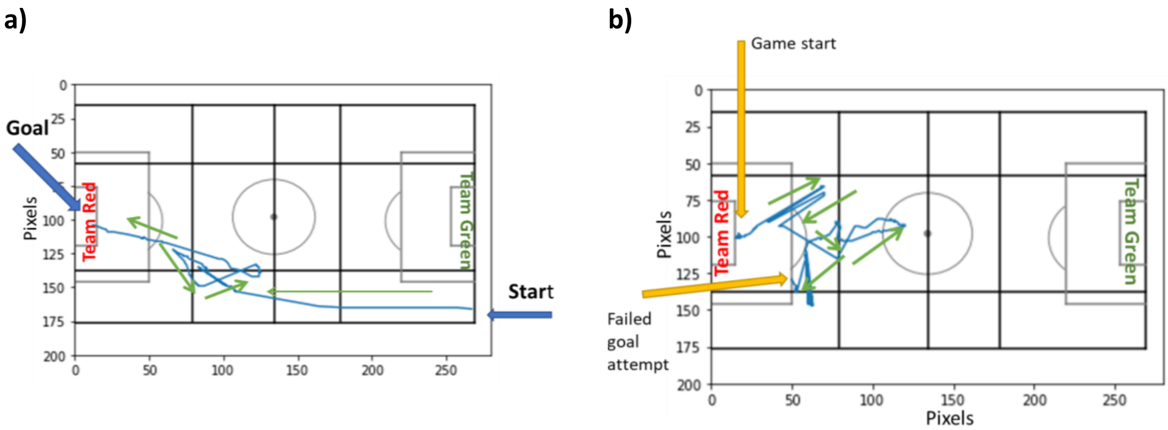


Figure 55 Goal attempts by the green team a) successful goal b) unsuccessful attempt

7. Discussion

There are many challenges during the ball and player detection such as the low resolution of the video making it difficult to detect the objects and also more probability of false positives. The penalty mark is another obstacle in the detection of the ball because it is also easily misclassified as the ball because of its shape and color. The color-based and template-based matching models mostly faced the problem of false detection with the penalty mark in the bigger field. Therefore, the color-based detection is performed only in the mini-map. Also, it is challenging to perform field modeling in the bigger field because of the ever-changing camera position and location. On the contrary, the mini-map is always completely visible and its location is also almost fixed on the screen so it can be used for accurate modeling of the field. The use of a mini-map for detection has a challenge because it keeps disappearing in the middle of the video so continuous tracking of the ball from the mini-map is not possible.

Understanding the location of the ball within the field can provide an interesting insight into the game strategy of the players. The action zone (figure 51) shows where the ball location is on the field during the first half of the game. It shows that the ball location in the first half was the mid-field but still it was more in Zone 1 (26%, red team defense area) in comparison to Zone 3 (20%). Also, the ball position by half (Figure 52) shows that the ball was mostly in the right half which is the green team's half, also it was team red who was mostly in the possession of the ball and the goal attempt is also more by the red team. All of these statistics show that the red team was the aggressive one while team green is more in the defense mode. Another important statistic for understanding the player's strategy is related to the utilization of the field for attacking purposes. The attack zone (figure 53) shows that team green mostly chose the middle part of the field (~58%) to attack while the red team prefer the top part of the field (~60%) to attack the opponent.

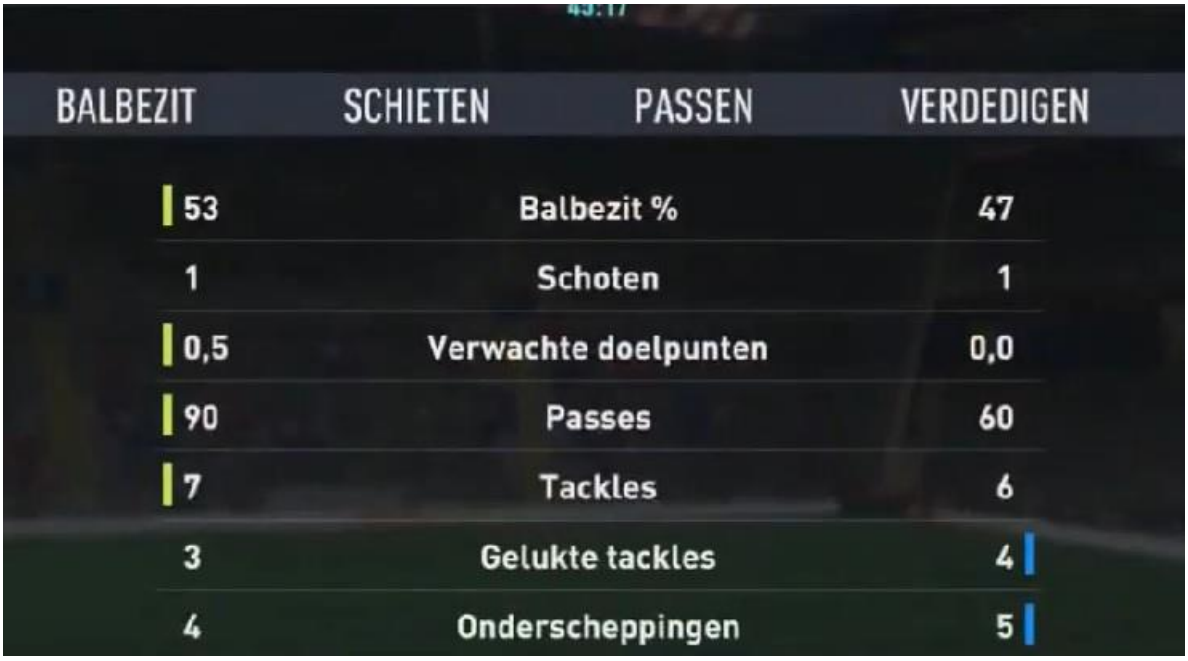


Figure 56 The scoreboard of the analyzed video after the first half.

Figure 56 shows the final statistics after the first half obtained from the video itself. According to this data, team red had 53% possession of the ball and team green had 47% possession of the ball. In the current study, the possession of the ball is detected on 10959 frames in the first half out of which on 6235 frames team red is in the possession of the ball which makes it ~ 57% possession of the ball while team green is detected to have the ball possession on 4724 frames i.e approximately 43%. These statistics are shown in figure 57. The model prediction is close to the values provided by the video scoreboard which shows the accuracy of the combined model.

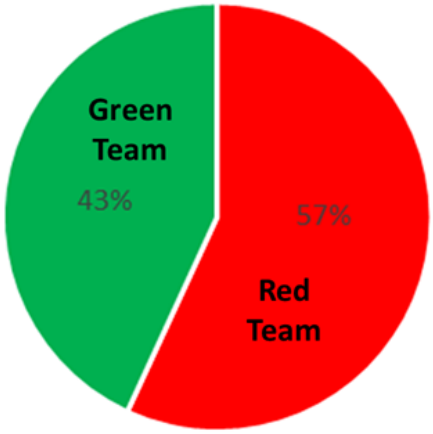


Figure 57 Ball possession by the teams

From the results of this study, it can be concluded that it is indeed possible to track and detect fast-moving objects such as balls with high precision (in the case of the Yolo-based model the precision of ball detection is ~ 97%). Furthermore, the extracted data can be used for game analysis. Still, there is ample room for improvement. The ball and player detection and tracking are not 100 % accurate because of the occlusion of the ball by players and also the player to player occlusion. The mitigation of this issue while analyzing a 2D video is extremely challenging and warrants further research. Another drawback of the model is that it does not detect all the players in a frame. It mostly detects players close to the ball. In eSports games, the player has control of only one player at a time which is usually close to the ball, therefore, the proposed model is still useful for the analysis of such types of games.

Since only 1 match with just the first half of the video is analyzed in this study, therefore, it is still difficult to formulate the player's strategy with confidence. To know more about the preference of field zone many other videos should be analyzed as the game tactics of the player can also differ from game to game. Though the proposed research focused on the eSports game, the developed model can also be used for the game analysis of real games up to a certain extent that does not have API availability such as local football games.

8. Conclusion

The FIFA 2022 eSports football games are drawing the interests of people at a rapid pace. This increase in interest motivates the more in-depth analysis of these games which has several similarities with the real football game except that these games are played by individual players with each other or with the AI. Since APIs are not available for these games therefore machine learning and computer vision based models have to be developed for this analysis.

To facilitate this analysis, two research questions are formulated in this thesis. The first research question "**h*ow to accurately detect and track ball and players in an eSport FIFA 2022 game video?*** " relates to the accurate detection and tracking of the ball and the players which is a challenging but highly important task. Three models are developed for this purpose. The first model which is based on the color-based detection provided too many false positives for the ball detection in the bigger field. This model can detect and track players on the mini-map with a precision of 88 %. Therefore, it is used for ball detection from the mini-map. The second method used is the template matching model which also provided a high number of false-positive results for both the ball and the players due to the continuously changing pose of these objects during the game. This model was not further used in this study.

The third model, YOLO-based ball and player detection provided results with a precision of 96 %. The main limitation of this model is that to detect the ball in the bigger field, the play-field modeling has to be performed which is quite challenging due to the movements of the camera position and location. Though this model is capable of slight color variation, it is still limited to 5 jersey colors. To make it general, a large annotated dataset is required and this is also the bigger drawback of this model as this requires a lot of time and effort. Training the YOLO model on more classes and data can take weeks of training time.

Considering the pros and cons of these models a combined model which couples the color-based ball detection on the mini-map and YOLO-based ball and player detection on the bigger field is created. The mini-map is used to track the ball because the entire field is visible in it all the time and also the location of the mini-map is almost fixed on the screen. The player's location while possessing the ball is determined from the bigger field using the YOLO model. When the ball and the player's possession of the ball are determined in the same frame, the ball location is tracked on the mini-map. With the provided precision numbers, it can be deduced that the developed model can accurately predict the ball and player movement, thus it answers the first research question.

The second research question **"h*ow the extracted data can be used to determine the game tactics and strategies of the player?*** " is related to the utilization of the data from the combined model for the game analysis to understand the tactics and strategies of the eSports players. For this purpose, several statistics such as the location of the ball in the field, attack and action zones for each team, and ball path for the goal attempt are calculated. According to the video analysis, the ball was in the green team's half which means that the red team was in the attacking mode for most of the time. This is true as they have made 4 goal attempts out of which 3 were successful. Although the green team has used the complete field for attacking, the most preferred location of the attack was the middle attack zone for this team. Similarly, for the red team, the preferred attacking area was the top attack zone. It is observed that the red team has used mostly the complete field for the goal attempt in comparison to the green team which mostly used the middle and lower attack zones of the field.

The results show that these types of games can be analyzed effectively using the developed model and it can an interesting insight into the game which answers the second research question of this thesis. To achieve its true potential, several games have to be analyzed which can be used to find the patterns of the eSport player's game. These patterns can provide insight into the tactics and strategies of the players which can be used for the prediction of their game.

9. Future Outlook

The combined model developed in this thesis can find the ball and player location quite accurately. Still, several shortcomings should be improved in future research on this topic. Since the data is annotated for 5 colors only, therefore the accuracy can be further increased by annotating more color images. The video analysis provided certain statistics related to the eSports player tactics and strategies. The true prediction of these attributes can happen only after analyzing several game videos which can provide more accurate patterns of the play thereby providing insight into the players' tactics. The mini-map keeps disappearing in the video which breaks the continuous detection and tracking of the ball. The utilization of the bigger field which comes with its complexity in terms of the field mapping complexity would provide more accuracy in this analysis. Furthermore, if only one model is used the code execution time can be reduced further. Even though the template matching does not provide useful results, more work can be done on this method as there is a possibility for the improvement of the model especially by providing multiple templates for one class (i.e. providing multiple templates for grey team players in different pose). However, due to time constraints, it was not done in this thesis and it could be interesting for future research on this topic.

References

[1] P. R. Kamble, A. G. Keskar, and K. M. Bhurchandi, “Ball tracking in sports: a survey,” *Artificial Intelligence Review*, vol. 52, no. 3, pp. 1655–1705, Oct. 2019, doi: 10.1007/s10462-017-9582-2.

[2] H. T. Chen, W. J. Tsai, S. Y. Lee, and J. Y. Yu, “Ball tracking and 3D trajectory approximation with applications to tactics analysis from single-camera volleyball sequences,” *Multimedia Tools and Applications*, vol. 60, no. 3, pp. 641–667, Oct. 2012, doi: 10.1007/s11042-011-0833-y.

[3] B. Vinoth Kumar, S. Abirami, R. J. Bharathi Lakshmi, R. Lohitha, and R. B. Udhaya, “Detection and Content Retrieval of Object in an Image using YOLO,” in *IOP Conference Series: Materials Science and Engineering*, Oct. 2019, vol. 590, no. 1. doi: 10.1088/1757-899X/590/1/012062.

[4] X.-F. Tong, H.-Q. Lu, and Q.-S. Liu, “An Effective and Fast Soccer Ball Detection and Tracking Method.”

[5] Z. Zhang, D. Xu, and M. Tan, “Visual measurement and prediction of ball trajectory for table tennis Robot,” *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 12, pp. 3195–3205, Dec. 2010, doi: 10.1109/TIM.2010.2047128.

[6] I. and K. Y. and O. Y. Ishii Norihiro and Kitahara, “3D Tracking of a Soccer Ball Using Two Synchronized Cameras,” in *Advances in Multimedia Information Processing – PCM 2007*, 2007, pp. 196–205.

[7] J. Liu, Z. Fang, K. Zhang, and M. Tan, “Improved high-speed vision system for table tennis robot,” in *2014 IEEE International Conference on Mechatronics and Automation, IEEE ICMA 2014*, 2014, pp. 652–657. doi: 10.1109/ICMA.2014.6885774.

[8] N. v Hoyningen-Huene and M. Beetz, “RAO-BLACKWELLIZED RESAMPLING PARTICLE FILTER FOR REAL-TIME PLAYER TRACKING IN SPORTS.”

[9] T. D’Orazio *et al.*, “An investigation into the feasibility of real-time soccer offside detection from a multiple camera system,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 12, pp. 1804–1818, Dec. 2009, doi: 10.1109/TCSVT.2009.2026817.

[10] P. Figueroa, N. Leite, R. M. L. Barros, I. Cohen, and G. Medioni, “Tracking soccer players using the graph representation.”

[11] X. Yu, X. Yan, T. sen Hay, and H. W. Leong, “3D Reconstruction and Enrichment of Broadcast Soccer Video,” 2004.

[12] A. Kumar, P. Shashidhar Chavan, S. David, P. Kelly, and N. E. O, “3D Estimation and Visualization of Motion in a Multicamera Network for Sports.”

[13] M. Barnard and J.-M. Odobez, “Robust playfield segmentation using MAP adaptation,” in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, 2004, vol. 3, pp. 610-613 Vol.3. doi: 10.1109/ICPR.2004.1334603.

[14] J.-L. Gauvain and C.-H. Lee, “Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains,” *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 2, pp. 291–298, 1994, doi: 10.1109/89.279278.

[15] A. Dearden, Y. Demiris, and O. Grau, “Tracking Football Player Movement From a Single Moving Camera Using Particle Filters.” [Online]. Available: http://www.bbc.co.uk/rd/projects/virtual/piero/index.shtml

[16] W. C. Naidoo and J. R. Tapamo, “Soccer Video Analysis by Ball, Player and Referee Tracking.” [Online]. Available: www.FIFA.com

[17] J. Ren, J. Orwell, G. A. Jones, and M. Xu, “Real-Time Modeling of 3-D Soccer Ball Trajectories From Multiple Fixed Cameras,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 3, pp. 350–362, 2008, doi: 10.1109/TCSVT.2008.918276.

[18] A. H. Mazinan and A. Amir-Latifi, “Applying mean shift, motion information and Kalman filtering approaches to object tracking,” *ISA Transactions*, vol. 51, no. 3, pp. 485–497, 2012, doi: 10.1016/j.isatra.2012.02.002.

[19] R. van der Merwe, N. Ando De Freitas, A. Doucet, and E. Wan, “The Unscented Particle Filter.” [Online]. Available: http://www.cs.berkeley.edurjfgf

[20] J.-Y. Kim and T.-Y. Kim, “Soccer Ball Tracking Using Dynamic Kalman Filter with Velocity Control,” in *2009 Sixth International Conference on Computer Graphics, Imaging and Visualization*, 2009, pp. 367–374. doi: 10.1109/CGIV.2009.87.

[21] P. R. Kamble, A. G. Keskar, and K. M. Bhurchandi, “A deep learning ball tracking system in soccer videos,” *Opto-Electronics Review*, vol. 27, no. 1, pp. 58–69, Mar. 2019, doi: 10.1016/J.OPELRE.2019.02.003.

[22] Y.-C. Huang, I.-N. Liao, C.-H. Chen, T. Ts`ı-Uí, ˙ Ik, and W.-C. Peng, “TrackNet: A Deep Learning Network for Tracking High-speed and Tiny Objects in Sports Applications*; TrackNet: A Deep Learning Network for Tracking High-speed and Tiny Objects in Sports Applications*,” *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2019, Accessed: Sep. 20, 2021. [Online]. Available: https://nol.cs.nctu.edu.tw:234/open-source/TrackNet/.

[23] M. Buric, M. Pobar, and M. Ivasic-Kos, “Ball detection using yolo and mask R-CNN,” 2018. doi: 10.1109/CSCI46756.2018.00068.

[24] L. Jiao, F. Zhang, F. Liu, S. Member, S. Yang, and Z. Feng, “A Survey of Deep Learning-Based Object Detection”, doi: 10.1109/ACCESS.2019.2939201.

[25] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Dec. 2016, vol. 2016-December, pp. 779–788. doi: 10.1109/CVPR.2016.91.

[26] D. and E. D. and S. C. and R. S. and F. C.-Y. and B. A. C. Liu Wei and Anguelov, “SSD: Single Shot MultiBox Detector,” in *Computer Vision – ECCV 2016*, 2016, pp. 21–37.

[27] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Transactions on Pattern Analysis and*

Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[28] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.0575

[29] M. and B. S. and H. J. and P. P. and R. D. and D. P. and Z. C. L. Lin Tsung-Yi and Maire, “Microsoft COCO: Common Objects in Context,” in *Computer Vision – ECCV 2014*, 2014, pp. 740–755.

[30] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite.” [Online]. Available: www.cvlibs.net/datasets/kitti

[31] P. Dollar, C. Wojek, B. Schiele, and P. Perona, “Pedestrian Detection: An Evaluation of the State of the Art,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 743–761, 2012, doi: 10.1109/TPAMI.2011.155.

[32] C. Guntuboina, A. Porwal, P. Jain, and H. Shingrakhia, “Video Summarization for Multiple Sports Using Deep Learning,” *ELCVIA Electronic Letters on Computer Vision and Image Analysis*, vol. 20, no. 1, pp. 99–116, May 2021, doi: 10.5565/rev/elcvia.1286.

[33] J. and Š. F. and M. J. Rozumnyi Denys and Kotera, “Non-causal Tracking by Deblatting,” in *Pattern Recognition*, 2019, pp. 122–135.

[34] J. Kotera, D. Rozumnyi, F. F. Filipšroubek, and J. Matas, “Intra-Frame Object Tracking by Deblatting.”

[35] J. Kotera and F. roubek, “Motion Estimation and Deblurring of Fast Moving Objects,” *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 2860–2864, 2018.

[36] A. Zita and F. Sroubek, “Learning-based Tracking of Fast Moving Objects,” May 2020, [Online]. Available: http://arxiv.org/abs/2005.01802

[37] X. F. Tong, H. Q. Lu, and Q. S. Liu, “An effective and fast soccer ball detection and tracking method,” in *Proceedings - International Conference on Pattern Recognition*, 2004, vol. 4, pp. 795–798. doi: 10.1109/ICPR.2004.1333892.

[38] D. Liang, Y. Liu, Q. Huang, and W. Gao, “A scheme for ball detection and tracking in broadcast soccer video,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2005, vol. 3767 LNCS, pp. 864–875. doi: 10.1007/11581772_76.

[39] G. Welch and G. Bishop, “An Introduction to the Kalman Filter.” [Online]. Available: http://www.cs.unc.edu/~gb

[40] X. Yu, Q. Tian, and K. W. Wan, “A novel ball detection framework for real soccer video,” in *Proceedings - IEEE International Conference on Multimedia and Expo*, 2003, vol. 2. doi: 10.1109/ICME.2003.1221604.

[41] R. Brunelli~’ and T. Poggio$, “TEMPLATE MATCHING: MATCHED SPATIAL FILTERS AND BEYOND,” 1997.

[42] K. C. A. Kumar, P. Parisot, and C. de Vleeschouwer, “Demo: Spatio-temporal template matching for ball detection,” in *2011 Fifth ACM/IEEE International Conference on Distributed Smart Cameras*, 2011, pp. 1–2. doi: 10.1109/ICDSC.2011.6042940.

[43] G. Feldman, “A 3×3 isotropic gradient operator for image processing pv-A Tool for Color Segmentation and Palette Visualization View project Curved Displays for Interactive Teleconferencing View project Irwin Sobel Consultant (HP Labs Retired-8Mar13),” 2018. [Online]. Available: https://www.researchgate.net/publication/285159837

[44] S. H. Khatoonabadi and M. Rahmati, “Automatic soccer players tracking in goal scenes by camera motion elimination,” *Image and Vision Computing*, vol. 27, no. 4, pp. 469–479, Mar. 2009, doi: 10.1016/j.imavis.2008.06.015.

[45] D. Acuna, “Towards Real-Time Detection and Tracking of Basketball Players using Deep Neural Networks.” [Online]. Available: http://goo.gl/ZPWStU

[46] W.-L. Lu, J.-A. Ting, J. J. Little, and K. P. Murphy, “IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE 1 Learning to Track and Identify Players from Broadcast Sports Videos.”

[47] J. Liu, X. Tong, W. Li, T. Wang, Y. Zhang, and H. Wang, “Automatic player detection, labeling and tracking in broadcast soccer video,” *Pattern Recognition Letters*, vol. 30, no. 2, pp. 103–113, Jan. 2009, doi: 10.1016/j.patrec.2008.02.011.

[48] M. Herrmann, M. Hoernig, and B. Radig, “Online Multi-player Tracking in Monocular Soccer Videos,” *AASRI Procedia*, vol. 8, pp. 30–37, 2014, doi: 10.1016/j.aasri.2014.08.006.

[49] D. Acuna, “Towards Real-Time Detection and Tracking of Basketball Players using Deep Neural Networks.” [Online]. Available: http://goo.gl/ZPWStU

[50] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, 2005, vol. I, pp. 886–893. doi: 10.1109/CVPR.2005.177.

[51] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, 2005, vol. 1, pp. 886–893 vol. 1. doi: 10.1109/CVPR.2005.177.

[52] B. Han and L. S. Davis, “Density-Based Multifeature Background Subtraction with Support Vector Machine,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, pp. 1017–1023, 2012, doi: 10.1109/TPAMI.2011.243.

[53] S. Hurault, C. Ballester, and G. Haro, “Self-Supervised Small Soccer Player Detection and Tracking,” in *MMSports 2020 - Proceedings of the 3rd International Workshop on Multimedia Content Analysis in Sports*, Oct. 2020, pp. 9–18. doi: 10.1145/3422844.3423054.

[54] G. Zhu, C. Xu, Q. Huang, and W. Gao, “Automatic Multi-Player Detection and Tracking in Broadcast Sports Video using Support Vector Machine and Particle Filter,” in *2006 IEEE International Conference on Multimedia and Expo*, 2006, pp. 1629–1632. doi: 10.1109/ICME.2006.262859.

[55] Y. Yang and D. Li, “Robust player detection and tracking in broadcast soccer video based on enhanced particle filter,” *Journal of Visual Communication and Image Representation*, vol. 46, pp. 81–94, Jul. 2017, doi: 10.1016/j.jvcir.2017.03.008.

[56] Y. Zhang, Z. Chen, and B. Wei, “A Sport Athlete Object Tracking Based on Deep Sort and Yolo V4 in Case of Camera Movement,” in *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, 2020, pp. 1312–1316. doi: 10.1109/ICCC51575.2020.9345010.

[57] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.10934

[58] B. Thulasya Naik, M. Farukh Hashmi, C. Author, and M. Farukh Hashmi mdfarukh, “Ball and Player Detection & Tracking in Soccer Videos Using Improved YOLOV3 Model,” 2021, doi: 10.21203/rs.3.rs-438886/v1.

[59] Y. Yoon *et al.*, “Analyzing Basketball Movements and Pass Relationships Using Realtime Object Tracking Techniques Based on Deep Learning,” *IEEE Access*, vol. 7, pp. 56564–56576, 2019, doi: 10.1109/ACCESS.2019.2913953.

[60] A. Alahi, Y. Boursier, L. Jacques, and P. Vandergheynst, “Sport players detection and tracking with a mixed network of planar and omnidirectional cameras,” in *2009 Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*, 2009, pp. 1–8. doi: 10.1109/ICDSC.2009.5289406.

[61] M. Buric, M. Ivasic-Kos, and M. Pobar, “Player Tracking in Sports Videos,” in *2019 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, 2019, pp. 334–340. doi: 10.1109/CloudCom.2019.00058.

[62] L. Sun and G. Liu, “Field lines and players detection and recognition in soccer video,” in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 1237–1240. doi: 10.1109/ICASSP.2009.4959814.

[63] A. Ekin and A. Murat Tekalp, “Robust Dominant Color Region Detection with Applications to Sports Video Analysis.”

[64] M. Luo, Y. F. Ma, and H. J. Zhang, “Pyramidwise structuring for soccer highlight extraction,” in *ICICS-PCM 2003 - Proceedings of the 2003 Joint Conference of the 4th International Conference on Information, Communications and Signal Processing and 4th Pacific-Rim Conference on Multimedia*, 2003, vol. 2, pp. 945–949. doi: 10.1109/ICICS.2003.1292597.

[65] B. Li and I. Sezan, “Event Detection and Summarization in American Football Broadcast Video *,” 2002. [Online]. Available: http://proceedings.spiedigitallibrary.org/

[66] D. Zhong and S. F. Chang, “Structure analysis of sports video using domain models,” in *Proceedings - IEEE International Conference on Multimedia and Expo*, 2001, pp. 713–716. doi: 10.1109/ICME.2001.1237820.

[67] M. Petkovic, W. Jonker, and Z. Zivkovic, “Recognizing Strokes in Tennis Videos Using Hidden Markov Models.”

[68] Y. Liu, W.-Y. Ma, and H.-J. Zhang, “Effective Feature Extraction for Play Detection in American Football Video,” 2005.

[69] K. CHOI, B. PARK, S. LEE, and Y. SEO, “TRACKING THE BALL AND PLAYERS FROM MULTIPLE FOOTBALL VIDEOS,” *International Journal of Information Acquisition*, vol. 03, no. 02, pp. 121–129, Jun. 2006, doi: 10.1142/s0219878906000897.

[70] “https://datacarpentry.org/image-processing/06-blurring/.”

[71] B. Hdioud, M. el Haj Tirari, R. Oulad Haj Thami, and R. Faizi, “Detecting and shadows in the HSV color space using dynamic thresholds,” *Bulletin of Electrical Engineering and Informatics*, vol. 7, no. 1, pp. 70–79, Mar. 2018, doi: 10.11591/eei.v7i1.893.

[72] Dilation, “https://nl.mathworks.com/ help/images/morphological-dilation-and-erosion.html.”

[73] “Computer Vision - A Modern Approach - D. Forsyth, J. Ponce”.

[74] S. Suzuki, “Topological Structural Analysis of Digitized Binary Images by Border Following,” 1985.

[75] I. Journal, H. A. Patel, and D. G. Thakore, “Moving Object Tracking Using Kalman Filter Related papers International Journal of Computer Science and Mobile Computing Moving Object Tracking Using Kalman Filter,” 2013. [Online]. Available: www.ijcsmc.com

[76] G. Welch and G. Bishop, “An Introduction to the Kalman Filter.” [Online]. Available: http://www.cs.unc.edu/~gb

[77] J. Y. Kim and T. Y. Kim, “Soccer ball tracking using dynamic kalman filter with velocity control,” in *Proceedings of the 2009 6th International Conference on Computer Graphics, Imaging and Visualization: New Advances and Trends, CGIV2009*, 2009, pp. 367–374. doi: 10.1109/CGIV.2009.87.

[78] T. Mahalakshmi, R. Muthaiah, and P. Swaminathan, “Review Article: An Overview of Template Matching Technique in Image Processing,” *Research Journal of Applied Sciences, Engineering and Technology*, vol. 4, no. 24, pp. 5469–5473, 2012.

[79] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection.” [Online]. Available: http://pjreddie.com/yolo/

[80] D. Nah;D Ebrnh, ; Majd, and X. Ye, “A-~ • , THE PROJECT HAS BEEN ACCEPTED BY THE PROJECT COMMITTEE IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF.”

[81] C. Kanan and G. W. Cottrell, “Color-to-grayscale: Does the method matter in image recognition?,” *PLoS ONE*, vol. 7, no. 1, Jan. 2012, doi: 10.1371/journal.pone.0029740.

[82] “https://towardsdatascience.com/canny-edge-detection-step-by-step-in-python-computer-vision-b49c3a2d8123.”

[83] F. Öberg, “Football analysis using machine learning and computer vision.”

[84] M. Abo-Zahhad, R. R. Gharieb, S. M. Ahmed, and A. A. E.-B. Donkol, “Edge Detection with a Preprocessing Approach,” *Journal of Signal and Information Processing*, vol. 05, no. 04, pp. 123–134, 2014, doi: 10.4236/jsip.2014.54015.

[85] D. G. Falconer, "TARGET TRACKING WITH THE HOUGH TRANSFORM.," pp. 249–252, 1977, doi: 10.1109/acssc.1977.748926.