Analysis of player tracking data extracted from football match feed

Swetha SASEENDRAN, Sathish Prasad Vetrivel THANALAKSHMI,
Swetha PRABAKARAN, Priyadharsini RAVISANKAR

Department of Computer Science Engineering
Sri Sivasubramaniya Nadar College of Engineering, India

Abstract: Data analytics and AI have become extremely relevant in today's football landscape. Data is benefiting clubs in gaining a competitive advantage on and off the field by empowering them to harvest information for improving player performance, decreasing injuries, and increasing commercial efficiency. Tracking data in football, that is the data of x, y coordinates of all 22 players on the pitch every second adds a lot of value in terms of a team's decision-making and recruitment strategy. Via the use of this data, each player's decision-making ability is also measured. The players and the ball in each frame are identified using YOLOv5, which returns their coordinates. These detections are then passed to DeepSORT, which assigns IDs to each player and keeps track of the frame by frame, by feeding each player detection to the model. The K-Means model is employed to determine the jersey color of the players to identify the two teams. Finally, the detected coordinates are multiplied with the homogeneous matrix computed using the Sports Camera Calibration through the Synthetic Data paper approach to accomplish the perspective transform. The mathematical model hypothesized and implemented uses pitch control and expected threat to assess each player's decision-making ability, which will be a strive to enhance the recruiting of players.

Keywords: Computer Vision, Perspective Transformation, YOLOv5, DeepSORT, Object Localisation, Object tracking, Data Analytics, Pitch Control, Expected Threat, Voronoi Diagram

1. Introduction

Football has been a very complex sport in terms of understanding as well as to be explained. Understanding the context and contextualization of football data helps us to well acknowledge the game flow and the players individually. The link between sports and computer science is well entrenched now. Putting the data, technology, and football all together points out innovation in football has resulted in the creation of decision-making evaluation of the players among other game analysis systems, where these systems analyze a player's critical moves individually in order to evaluate the decision-making ability of them.

We propose that the collaboration between sports and computer science should work in close cooperation as in teams and combine the approaches of both the domains rather than working in a segregated fashion. Having applied techniques from computer science to sports research designs, sport scientists could arrive at performance from various perspectives and a more in-depth understanding of their research questions. Tracking data is so niche that there is only one publicly available data of a match for research purposes. Understanding the context of this data helps us to well acknowledge the gameplay of the players individually. This technology today is only accessible to big-time football clubs and teams with an overabundance of funds. This is because extracting tracking data costs about 60,000 pounds for a team in one season. The idea is to use computer vision techniques to extract tracking data for little to almost no cost, so that this technology will be accessible to small scale football companies and local teams.

2. Literature survey

A player detection model by (Rangappa, Li & Qian, 2021) can be customized to allow it to identify people in video clips. It can also be used to track and identify soccer players. He also presents a method for assessing the performance of the model. The paper talks about converting Digits to Numbers, which is not a reliable method for long-term tracking, as it would not work with the match feed since the numbers would be too pixelated. Further, these models can be used for any other sporting video to generate tracking data as their scope can be widened.

https://doi.org/10.33436/v33i2y202307

There are various image processing techniques and methods that can be used to track a person (Pers & Kovacic, 2000). In certain cases, such as when the camera is not calibrated properly, the automatic tracking process must be interrupted manually. This paper was only used to test the system in a soccer match. Although the paper talks about the model's potential applications in other sports, its implementation is still not yet widely explored.

A method for locating a sports field based on a single game broadcast image was proposed by (Homayounfar et al., 2017). The problem of field localization is framed as a Markov Random Field inference with potentials derived from a deep semantic segmentation network. This method is totally automated and relies just on a single image from the game's broadcast footage. Soccer and hockey have been used to test this strategy. However, it ignores temporal information, resulting in some inconsistencies.

STN-Homography (Zhou & Li, 2019) is a neural network based on the spatial transformer network that immediately predicts the normalised homography matrix of two images. Hierarchical STN-Homography and sequence STN-Homography models are used to reduce homography estimation error. Experiments using the MSCOCO dataset show that the current research and development work.

Common object detection designs, as well as some tweaks and important tips for improving detection performance was focused in (Wu, 2018). It provides an overview of popular object detection algorithms as well as a performance comparison. This paper presents a comprehensive overview of deep learning object identification methods that address a variety of sub-problems such as obscurity, clutter, and resolution, to name a few.

Object detection with deep learning techniques (Zhao et al., 2019) focuses on the development of deep learning tools and its high-level features thereby addressing the problems in the existing features and their behavior towards neural architectures. It majorly talks about CNN, a representative tool of deep learning, and about object detection architectures, as well as their adaptations and relevant tips for improving detection performance. This paper surveys face detection and object detection by providing us an experimental analysis by comparing various methods and draws conclusions out of it. However, it doesn't give us a clear view to address the problems in the existing features and the detection techniques need to be improved further.

The ways to detect and track the moving objects using algorithms based on computer vision technology was proposed (Chen & Li, 2021). Here they first compare both the theoretical and experimental aspects to the background difference model method and detect the performance. And Robert edge detection operator is used to perform the edge detection of the image. Even though the algorithm has high efficiency, the accuracy doesn't meet the requirements fully and is yet to be improved.

A simple fusion of well-known techniques including the Kalman filter and the Hungarian algorithm was used (Bewley et al., 2016). But doesn't integrate the information of the image and its appearance to improve the performance like DeepSORT which uses deep association metrics. They mainly focused on reducing the number of identity switches. They have also established extensions and measurement-to-track association to reduce the identity switches at high frame rates. Time complexity and performance efficiency can be developed further to use the extensions in real time tracking.

An approach (Yang et al., 2021) in which texture and HSV aspects of a footballer's posture picture are extracted individually, then a dual-channel CNN is built and integrated with it. Finally, the acquired findings are fed into a fully connected CNN, which estimates and constructs the footballer's posture image. This study uses a huge set of data to undertake experimental testing and comparison analysis, and it indicates that the suggested algorithm's experimental analysis delivers higher performance outcomes.

A ball-passing network (Wang et al., 2020) to enable football teamwork analysis and present a new model for quality assessment based on times at its apex to measure team performance was developed. They also aimed to establish a third model to quantify the conversion flow of both teams and their approaches by presenting a comprehensive evaluation of the whole players. The main procedure and how the analysis is performed here are understood, however it only works on small size data, enabling the huge collection to be examined and improved.

Enhanced possibilities of pitch control model where the paper tries to analyze Wide open spaces in football was studied (Fernandez & Bornn, 2018). With only 3% of the time where a player is involved in the game, the remaining off the ball movements made to create spaces are studied in this paper. However, this does not evaluate pass values but rather focuses on the movements of players.

The first evolution in tracking data from using Voronoi (Spearman, 2016). In this paper physical metrics such as velocity of player, acceleration, speed of ball and player was considered to calculate area of control of a player in the pitch. Area of control was studied in this paper and nothing related to quantifying a pass was studied. We will use this as the base to calculate the area of control in a given instance. The pass success probability by considering the location of players, their speed, speed of the ball and time to control the ball was studied (Spearman, 2018). This was an evolution from quantifying pitch control. All these factors were considered to calculate pitch control in each zone for decision making.

A new system (Spearman et al., 2017) was proposed where quantifying goals should be more than just calculating expected goals model. With the help of tracking data, the paper delves deeper into frames before a goal moves and tries to analyze patterns and see how players create spaces and add value to the team. However, this paper focuses on off-the-ball movements and doesn't investigate anything related to passes.

Individual movement models from positional data and demonstrating how to convert these estimates into precise and realistic control zones was estimated (Brefeld et al., 2019). Their method takes into consideration player characteristics, scales with massive quantities of data, and can be calculated efficiently in a distributed manner. However, while this work concentrates on off-the-ball movements and does not investigate passes, we will use their inputs to quantify the pitch control model.

3. System design

The architecture consists of the following modules: Model Creation which aims to create a computer vision model for extracting tracking data coordinates from the video input, that is the data of x, y coordinates of all 22 players on the pitch every frame. This includes Object Detection, Object Tracking, Jersey Color Detection and Perspective Transformation. Data Analysis where with the extracted tracking data, the intention is to evaluate and quantify the player's decision-making process and other inferences about the game that helps to improve the team's game strategy. Figure 1 shows the detailed architecture along with the models and algorithms used in this project.

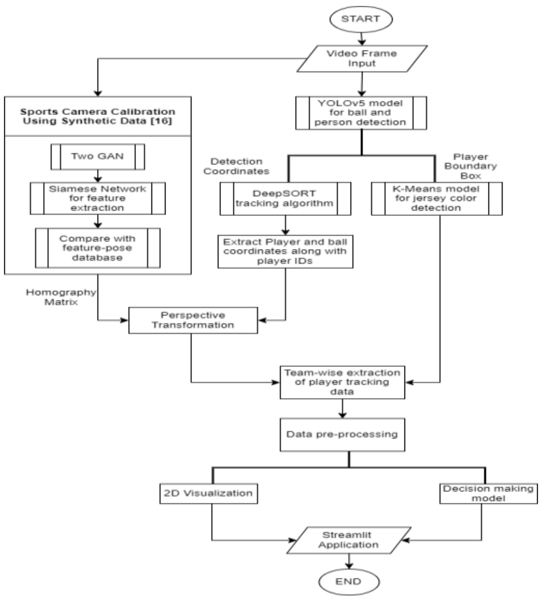

Figure 1. Overall System Architecture

3.1. Player detection

Player detection is one of the problems to solve when it comes to generating team/player statistics. To solve this problem using a deep learning approach made more sense as it requires a robust solution. Players are identified to get their positional coordinates from a given frame in order to pass them to the trackerto track them from frame to frame.

Pre-trained deep learning model is used for person detection. YOLOv5 has been used to plot the bounding boxes for detected players and sports-ball. The expression 'You Only Look Once' is abbreviated as YOLO. This is an algorithm for accurately identifying different objects in an image. Object detection in YOLO is done as a regression problem, and the identified images' class probabilities are provided (Redmon & Farhadi, 2017).

Four models were reviewed and YOLOv5 was chosen for object detection. **YOLOv3** meets the real-time requirements. But, nearby objects and smaller objects, on the other hand, are difficult to detect. The positioning accuracy of the object by YOLOv3 is low, and the false detection and miss detection is unavoidable. With various bags of tricks and modules, **YOLOv4** is a one-stage object identification approach that improves on YOLOv3. When objects in the image reveal unusual aspects of ratio, nevertheless, it does not generalise effectively. It also has a slower interference speed. When employing multibox, a **single shot detector** (SSD) requires only one shot to detect several objects present in a picture like YOLO. However, training the data is cumbersome and time-consuming in this case. The network is too sluggish at inference time, and training occurs in numerous phases. **YOLOv5** model is quicker than any other model, and it performs better at recognising tiny objects. With minimal to no overlapping boxes, the results are also cleaner.

3.2. Player tracking

Object tracking is a technique in which a software algorithm takes a series of initial object detections and assigns each one a unique ID, then tracks the identified objects as they move across frames in a video. The aim is to track the player and assign unique IDs to them to identify them from one frame to another. Players are tracked for N frames until the next identification, enabling any tracking errors to be corrected. It runs for N frames, with the detection task coming first, and each bounding box being given an ID at the conclusion of the detection step. The N value in the present implementation is 5.

DeepSORT has been utilized to track players and assign unique IDs to them. It is an improved version of SORT and works using Kalman Filter. Figure 2 shows the DeepSORT architecture.

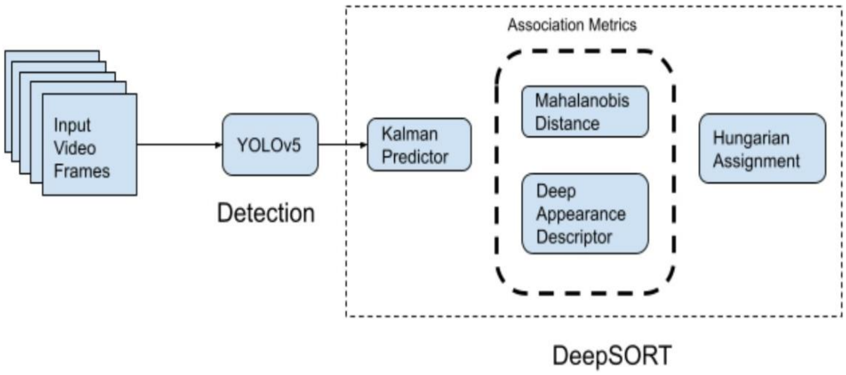

Figure 2. DeepSORT Architecture

The Kalman Filter is based on a uniform speed model and a gaussian distribution. When the object is visible, you hinge more on the sensor data and give it more value. When it is partially obscured, you rely on both motion and sensor information. If it's completely obstructed, you'll pay a lot of attention to motion data. The Kalman filter uses current readings to update predictions, thus anticipating its new location. The homography matrix is generated using the Intersection-Over-

Union (IOU) value across each identification and all bounding boxes from the current predictions. This is known as the Hungarian Algorithm and is useful when one object occludes another. DeepSORT introduces a new distance measure depending on the appearance of the object. A classifier that is trained carefully until it achieves a satisfactory degree of accuracy. The network's final classification layer is then removed, leaving a dense layer that gives a single feature vector ready for classification. This feature vector is referred to as an appearance descriptor. Following the acquisition of the appearance descriptor, closest neighbor queries are performed in the visual appearance to construct the Measurement-to-Track Association (MTA). The Mahalanobis players Distance is now used for MTA. (Wojke et al., 2017)

The following is the result obtained on implementing DeepSORT tracking algorithm with YOLOv5 object detection model. Figure 3 depicts a frame where the players are being detected by YOLOv5 model and tracked by DeepSORT.

Figure 3. Example output image for player, ball and Team Tracking using DeepSORT and YOLOv5

Other algorithms were surveyed before choosing DeepSORT. The straight-line distance between two places in Euclidean space is known as the Euclidean distance. The coordinates of the object are compared with the current and past frames in object tracking, and the distance between the current and prior frames is calculated using the Euclidean distance. This is the same object if the distance between two object values is less than the threshold; otherwise, a new ID is produced and assigned. The crucial point here is that the Euclidean standard assigns equal weight to all directions. Also, for the Euclidean distance to be meaningful, the objects must have non-zero characteristics, as it fails to compute the distance of items that are further away and only considers the distance of objects in the immediate vicinity, it eventually leads to errors in the tracking data's conclusions. It also has a high computing overhead and fails to manage occlusion. SORT achieves a decent overall performance in terms of tracking precision and accuracy. However, it returns a relatively large number of identity switches and has a shortcoming in tracking through occlusions and changing views, among other things. DeepSORT, which is an extension of SORT, is the most popular and commonly used elegant object tracking system (Simple Real-time Tracker). It records not just distance and velocity, but also the person's appearance. It allows us to add this functionality by computing deep features for each bounding box and factoring in the tracking algorithm based on deep feature similarity. DeepSORT is the quickest of the bunch with its, having handled the occlusion efficiently. It averaged 16 frames per second while retaining high accuracy, making it an excellent choice for multiple item recognition and tracking.

3.3. Jersey color detection

To classify the two teams in play without the need for manual intervention, jersey color identification should be an automated procedure. The player detection from the previous section will be applied, and each player identified should be assigned to either Team A or Team B based on their jersey color. To eliminate inconsistencies, the referees must be manually marked. However, this only must be done once, as the model will track the referees as well.

K-means locating k centroids and then to each cluster data point is assigned to that centroid that is closest to it, resulting in as compact a set of centroids as feasible. Each image contains n x m pixels, which are each made up of three components: RGB. A value will be grouped for each pixel in the picture. As a result, comparable pixel values in terms of color are grouped together. Here are two classes that correspond to K-means. The most probable color to be identified is green, due to the green tint of the grass. The second color will be used as the jersey color, and it will be contrasted with the basic colors to determine the jersey color. Figure 4 shows colored boundary box based on the jersey color identified by the K-Means model.

Figure 4. Boundary Box colored based on the team identified by the K-Means model

3.4. Perspective transformation

Homography is a transformation that converts points in one plane to equivalent points in another. In a typical video broadcast there are different perspectives or camera views that are used to cover any given football game. The birds eye view is the most common as it covers a sizeable portion of the football pitch that is relevant at any given time. However, sometimes closeup views of some players/events/fans are more appealing to the viewers and hence those views are shown on broadcast. For a computer vision algorithm like ours to gather any information it is only possible from birds eye view frames as a sizeable portion of the pitch is covered from which to extract information. The translation of a three-dimensional world into a two-dimensional representation is called perspective transformation. This is done to determine the football field's boundaries so that we may transfer actual space to a 2D coordinate system and display the tracking data. A perspective transform, a matrix operation that projects a set of points from one 2D plane to another, can be used to accomplish this mapping. The detections may then be transformed from video to bird's-eye perspective using Perspective Transformation utilizing these reference points (Cheng et al., 2019).

GAN is made up of two models, the generator (G) and discriminator (D), that are trained simultaneously and pitted against each other. The generator model learns to produce likely data, whereas the discriminator model learns to separate genuine data from phoney data generated by G. As a result, the training objective function may be described as a two-player min–max game, with the generator G attempting to reduce the objective and the discriminator D attempting to maximise it. As a result, it's called adversarial. The Pix2Pix architecture is built on conditional GAN. It uses an input picture or other aiding information to condition the generator and discriminator, resulting in targeted image production from the target domain. As a result, cGANs are well-suited to image-to-image translation.

Estimation of Homography Matrix Module is a PyTorch implementation of **Sports Camera Calibration via Synthetic Data paper (**Chen & Little, 2019). There are four sub-modules in this:

- GAN pix2pix for separating the grass pitch from the rest of the stadium (Isola et al., 2017);
- pix2pix GAN network for detecting football field lines (Isola et al., 2017);
- Database of football field line images (Homayounfar et al., 2017).

A Siamese network for evaluating single shot similarity to get the homography matrix. (Utkin et al., 2021).

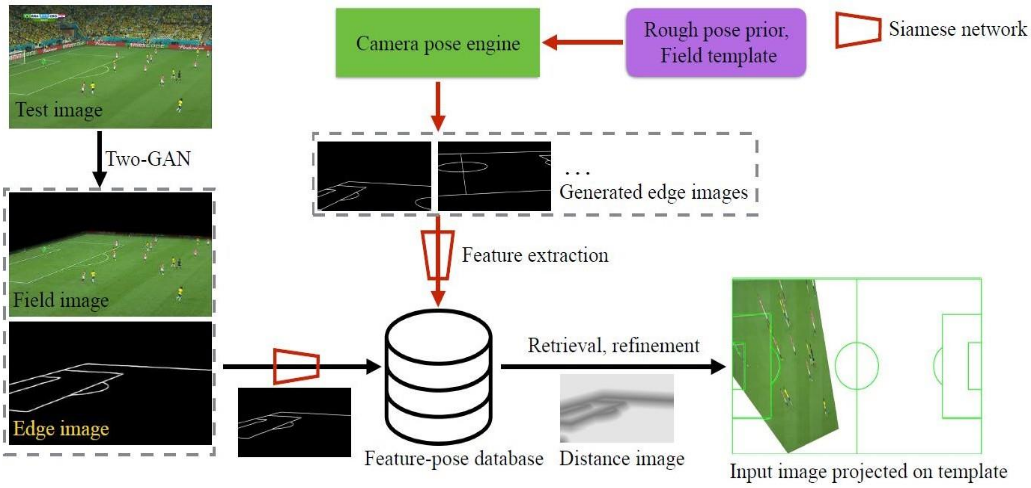

Figure 5. Sports camera calibration pipeline (Chen & Little, 2019)

Figure 5 shows the pipeline used in the Sports Camera Calibration using Synthetic Data. The World Cup dataset was utilized, which contains frames from 20 different football matches broadcast during the 2014 World Cup (Homayounfar et al., 2017; Pers & Kovacic, 2000). The camera view is projected into the bird's eye view using 210 frames from the training dataset and 200 frames from the test dataset, each with a homography matrix. The camera pose engine generates many cameras poses using the equation:

$$P = KQ_{\phi}Q_{\theta}S[I - C] \tag{1}$$

Re-parameterization of formula (1) is done by sampling the camera location C, tilt angle, and focal lens to get many cameras to pose examples, creating a homography dataset. As a conditional GAN, the Pix2Pix model learns a projection for y from an observed picture and a random noise vector. A basic GAN, on the other hand, learns projection for y from a random noise vector z. The generator G creates new data points, while the discriminator D examines them for legitimacy; that is, the discriminator determines whether each instance of data corresponds to the real training dataset. The court identification task is handled primarily by the perspective detection module, which facilitates the extraction of court characteristics. It makes use of two identical Pix2Pix (Isola et al., 2017) neural networks, which are conditional generative adversarial networks CGANs. In terms of training input, a conditional GAN varies from a basic GAN model. A CGAN's generator and discriminator are both conditioned on additional input data, whereas simple GANs are not.

The edge images are formed via a two-GAN network (two pix2pix GANs concatenated). (Chen & Little, 2019; Koeferl et al., 2020) emphasizes the notion of a dual CGAN, in which the first CGAN handles image segmentation and then feeds its result to the second CGAN, which conducts feature identification. Then, using the siamese network (Utkin et al., 2021), features correlated with each of the edge images are extracted and compared to extracted features recorded in the database during the training phase to discover the database's most similar edge image.

Using the Siamese network, various characteristics associated with each of the edge pictures are extracted and compared to the database to determine the homography matrix. This network will receive two images, possibly from the same or separate classes. The result will be a floating-point value between 0 and 1, with 1 signifying that the two images are from the very same class and 0 indicating that they are from different ones. There are multiple convolutional and pooling layers in a CNN model, accompanied by several dense layers and a softmax output layer. Aside from the lack of a softmax layer, siamese networks have a comparable convolutional and pooling layer arrangement to that of CNN. Because the network has two images as inputs, the model will have two dense layers, as aforementioned. Now, the difference between these two layers' outputs is computed and transmitted to a single sigmoid-activated neuron. Finally, by establishing that the relevant affine matrix was recorded in the database during the training phase and the football image was manipulated into itstop view.

The perspective transform is made possible by multiplying the tracked foot coordinates of each player to the homography matrix which will produce the 2D coordinates of the player in the template. Figure 6 represents the image template of the static top view of the court.

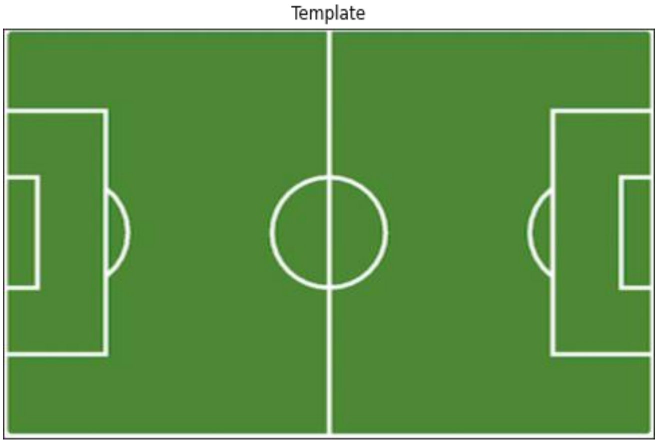

Figure 6. Static top view of the court template

Figure 7 shows the template being warped in the right as per the homography matrix extracted from the frame on the left. (x_1, y_1) and (x_2, y_2) are two sets of points and X_i is defined as [x_i, y_i, 1]⊤. The planar homography is defined as H that relates the transformation between the 2 planes generated by X_1 and X_2 as:

$$X_2 = HX_1 = \begin{matrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{matrix} X_1 \tag{2}$$

Figure 7. Warped Template from computed homography matrix

$$X_2 = \frac{h_{11}x_1 + h_{12}y_1 + h_{13}}{h_{31}x_1 + h_{32}y_1 + 1} \tag{3}$$

$$Y_2 = \frac{h_{21}x_1 + h_{22}y_1 + h_{23}}{h_{31}x_1 + h_{32}y_1 + 1} \tag{4}$$

Figure 8. Image shows the 2D visualisation of coordinates obtained using perspective transform

4. Quantifying decision making ability

Upon generating the x, and y coordinates the next step is to quantify the decision making ability of the players. To begin with, firs divide a football pitch into bins with 16x12 dimensions. Bins are basically a square on the field where the count of the number of involvements in that particular box and also to find it easy to assign values to that bin.

Expected threat (xT) bin values are used for this where xT means the amount of threat caused by the attacking team by moving the ball from one zone to another zone. Another concept that's being used is the pitch control matrix. Pitch control means given a bin on the field, what's the probability that a given team is said to retain possession. With the help of these two concepts, quantifying the Decision making of a player is performed. Event data is the type of data where that only has the coordinates of players who are involved on the ball and the positions of the remaining players are unknown. This is going to use tracking data to obtain the location of other players in the frame.

The linking column between both the tracking data and event data will be the frame ID. The columns in tracking data will contain only the coordinates, whereas the event data will contain the coordinates of the player involved and also the action performed.

So initially event data is considered and input from the user to choose the player whose decision making is going to be quantified. Then filter the action to successful passes. Now having the frame ID with which makes a way to retrieve the location of all players from the tracking data. The drawback of considering only Voronoi was that the only positions of players were considering and their acceleration and speed was neglected which is supposed to be a crucial factor in terms of deciding who is going to win the ball. Hence, Pitch control along with the xT is considered.

First and foremost, calculation of decision making value frame by frame is performed. Then, first identify which player decision making is have to be calculated and then filter all the frames in which the chosen player was the player who made the pass with the help of event data and tracking data. With the positions of 22 players on the pitch, xT values which are predefined are calculated first.

Now using existing model xT, values are assigned to each of the player locations. $xT_{x,y}$ is given by the equation:

$$xT_{x,y} = (s_{x,y}\ x\ g_{x,y}) + \left(m_{x,y} x \sum_{z=1}^{16} \sum_{w=1}^{12} T(x,y) \rightarrow (z,w)\, xT(z,w) \right) \tag{5}$$

- **Move probability** $m_{x,y}$: The probability that a player in possession of the ball in the zone (x,y) would choose to move the ball (i.e. pass or dribble) as their next action.
- **Shoot probability** $s_{x,y}$: The probability that a player in zone (x,y) who has possession of the ball would opt to shoot as their next action. When a player gets possession of the ball, his or her options are limitedto moving or shooting.
- **Move transition matrix** $T_{x,y}$ **:** The probability that a player will move the ball to each of the other zonesif they move the ball from zone (x,y).

- **Goal probability** $g_{x,y}$ **:** The probablity that if a player moves the ball from zone to zone, the ball will bemoved to each of the other zones is increased (x,y).

Then move on to calculate the pitch control value which is dependent on the positions of the players on the pitch. In order to calculate the pitch control, the positions of all players are considered and three basic questions are answered mathematically. On the given bin what would be the time taken for the ball to arrive? How long would it take for each player to arrive and what is the total probability that each team will control the ball after it has arrived? Now this calculation is repeated in every bin and allocate values which will be visualized later.

Since coordinates of all players are obtained, calculating their distance to each bin is easy, and to calculate the time they'll take to reach the first bin, then do make a few assumptions based on pre-set standards for football analytics. The first assumption is that the average ball speed is 15m/s, players have a maximum speed of 5m/s and their maximum acceleration is 7m/s/s and they also take the fastest possible path available. Now taking all these assumptions into account along with the x and y coordinates, find the probability of the above mentioned questions and assign it to every bin on the pitch.

Now 16x12 matrix with each bin having probability values is obtained and have a 16x12 matrix with pre- defined xT values. Now multiply these bin values to calculate the best passing option as positions with high expected threat and with high pitch control value will be the best possible option to pass on the pitch. Then to calculate the best passing option matrix for both initial frame and the end frame and the bin values of initial and final are subtracted which gives the value of the pass. The highest pass value is considered to be the best option in a given game situation. Now run a loop in all frames that the selected player was the passer and calculate his pass values and compare it with the highest pass value in each frame to calculate his decision-making ability.

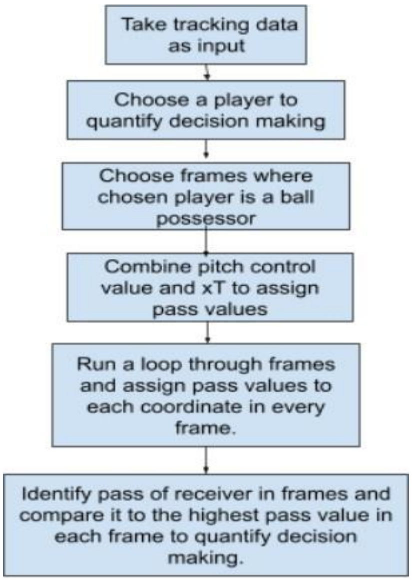

Figure 9. Calculating decision making

The formula to assign value to passes would be :

$$\textit{Pass Value} = (\textit{Pitch control value at final position} * \textit{xT at final position}) - (\textit{Pitch control value at initial position} * \textit{xT at the initial position}) \quad (6)$$

By assigning the above formula, the position of opponents and other factors are considered. By looping through the passes of a particular player and summing the pass value that the player claimed divided by the maximum value that he could have claimed decision making is quantified.

Figure 11 is a visualization of the 16x12 xT matrix. The darker the blue, the higher the threat if it passes into those areas of the pitch. Upon taking a particular frame from the tracking data and calculating the pitch control, Figure 10 is the visualization obtained. Here the red color indicates that if the ball is passed on the red colored space, the red team is more likely to retain possession and the same goes for the blue colored regions. Now with the pitch control value and the xT value, now multiply these two figures to identify possible passing options which hold high value.

Now decision making is quantified using the formula:

$$\textit{Decision Making} = \frac{\textit{Sum of pass value claimed}}{\textit{Sum of all maximum pass value the player could have claimed}} \quad (7)$$

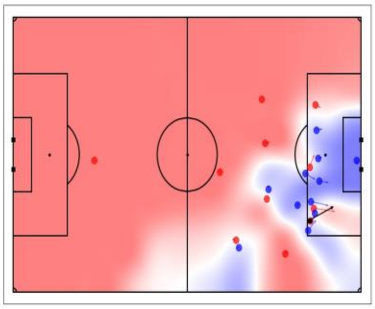

Figure 10. Pitch Control

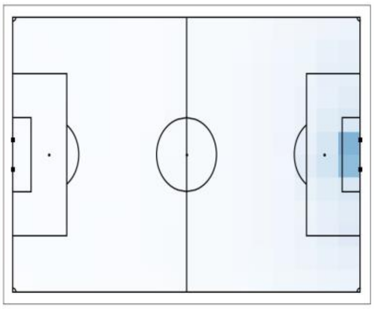

Figure 11. Expected Threat

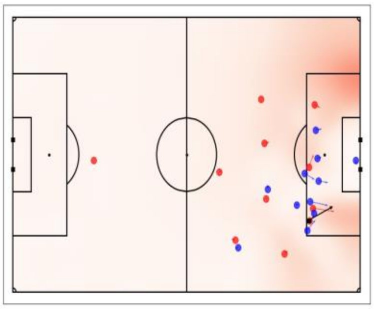

Figure 12. Pass Value

Figure 12 is a product of the expected threat matrix and the pitch control matrix. The greater the density of red, the better it's to pass to that particular location in the football field. xT matrix is independent of positions of other players on the pitch and pitch control does not help to identify good passing options. However, multiplying these two matrixes a new matrix is obtained which considers all the 22 players on pitch and the zones that are in our control. This way it is possible to quantify a pass and add a value to it.

REFERENCES

Bewley, A., Ge, Z., Ott, L., Ramos, F. & Upcroft, B. (2016) Simple online and real time tracking. In *2016 IEEE international conference on image processing (ICIP), 25-28 September 2016, Phoenix, AZ, USA*. pp. 3464-3468. IEEE. doi: 10.1109/ICIP.2016.7533003.

Brefeld, U., Lasek, J. & Mair, S. (2019) Probabilistic movement models and zones of control. *Machine Learning*. 108(1), 127-147.

Chen, C. & Li, D. (2021) Research on the detection and tracking algorithm of moving object in image based on computer vision technology. *Wireless Communications and Mobile Computing*. 2021, 1-7.

Chen, J. & Little, J. J. (2019) Sports camera calibration via synthetic data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. doi:10.48550/arXiv.1810.10658.

Cheng, S., Zhao, K. & Zhang, D. (2019) Abnormal water quality monitoring based on visual sensing of three-dimensional motion behavior of fish. *Symmetry*. 11(9), 1179.

Fernandez, J. & Bornn, L. (2018) Wide Open Spaces: A statistical technique for measuring space creation in professional soccer. In: *Sloan sports analytics conference*, 23-24 of February, Boston, USA.

Homayounfar, N., Fidler, S. & Urtasun, R. (2017) Sports field localization via deep structured models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 5212-5220.

Isola, P., Zhu, J.Y., Zhou, T. & Efros, A.A. (2017) Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 1125-1134.

Koeferl, F., Link, J. & Eskofier, B. (2020) Application of SORT on Multi-Object Tracking and Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*.

Pers, J. & Kovacic, S. (2000) Computer vision system for tracking players in sports games. In *IWISPA 2000. Proceedings of the First International Workshop on Image and Signal Processing and Analysis. In conjunction with 22nd International Conference on Information Technology Interfaces. IEEE*. pp. 177-182.

Rangappa, S., Li, B. & Qian, R. (2021) Tracking and identification for football video analysis using deep learning. In *Thirteenth International Conference on Machine Vision*, Vol. 11605. pp. 448-456. SPIE.

Redmon, J. & Farhadi, A. (2017) YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 7263-7271.

Spearman, W. (2016) Quantifying Pitch Control. *OptaPro Analytics Forum*.

Spearman, W. (2018) Beyond expected goals. In *Proceedings of the 12th MIT sloan sports analytics conference*. pp. 1-17.

Spearman, W., Basye, A., Dick, G., Hotovy, R. and Pop, P. (2017) Physics-based modeling of pass probabilities in soccer. In *Proceeding of the 11th MIT Sloan Sports Analytics Conference*.

Utkin, L., Kovalev, M. & Kasimov, E. (2021) An explanation method for siamese neural networks. In *Proceedings of International Scientific Conference on Telecommunications, Computing and Control: TELECCON 2019* pp. 219-230. Springer Singapore.

Wang, Y., Wang, H. & Qiu, M. (2020) Performance Analysis of Everton Football Club Based on Tracking Data. In *2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS)*. pp. 49-53.

Wojke, N., Bewley, A. & Paulus, D. (2017) Simple online and realtime tracking with a deep association metric. In *2017 IEEE international conference on image processing (ICIP)*. pp. 3645-3649.

Wu, J. (2018) Complexity and accuracy analysis of common artificial neural networks on pedestrian detection. In *MATEC Web of Conferences*, 232, (01003). EDP Sciences.

Yang, G., Wang, L., Xu, X. & Xia, J. (2021) Footballer action tracking and intervention using deep learning algorithm. *Journal of Healthcare Engineering*, *2021*, 1-9.

Zhao, Z. Q., Zheng, P., Xu, S. T. & Wu, X. (2019) Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 30(11). pp. 3212-3232.

Zhou, Q. & Li, X. (2019) STN-homography: Direct estimation of homography parameters for image pairs. *Applied Sciences*, 9(23). pp. 5187.

Swetha SASEENDRAN obtained her Bachelor's Degree in Computer Science and Engineering in 2022, at Sri Sivasubramaniya Nadar College of Engineering, Chennai, Tamil Nadu India. She is currently working as a Technology Analyst at Citibank India. Her areas of research interests include AI and Software Engineering.

Sathish Prasad Vetrivel THANALAKSHMI obtained his Bachelor's Degree in Computer Science and Engineering in 2022, at Sri Sivasubramaniya Nadar College of Engineering, Chennai, Tamil Nadu. He is currently a Data analyst at Ronnie Dog Media, Belgium. His areas of research interest includes Business Intelligence and Strategy, Operations Related Data Analytics.

Swetha PRABAKARAN obtained her Bachelor's Degree in Computer Science and Engineering in 2022, at Sri Sivasubramaniya Nadar College of Engineering, Chennai, Tamil Nadu. She is currently working as an Operational analyst at Gamenation, India. Her areas of research interests include Machine Learning and Web Development.

Priyadharsini RAVISANKAR is working as Associate Professor in department of Computer Science and Engineering, Sri Sivasubramaniya Nadar College of Engineering, Tamilnadu. She received her Ph.D in Information and Communication Engineering from Anna University. Her areas of research interest include underwater acoustic image processing, Computer Vision, and Data analytics.