A Survey on Object Detection and Tracking in Soccer Videos

Huda Dheyauldeen Najeeb [1*], Rana Fareed Ghani [2]

[1]*Department of Public Relations, College of Media, Al Iraqia University*

[2]*Department of Computer Science, College of Science, Technology University*

[*]*Corresponding Author:* 111806@student.uotechnology.edu.iq

Received 2/9/2020 , Accepted 21/10/2020 , published:25/10/2020

DOI: 10.52113/2/08.01.2021/1-13

Abstract: The players and ball are the most important object in soccer game videos and detected them are a challenging task because of many difficulties, such as shadow and illumination, ball size, several other objects look like a ball, often the ball overlapping with players or merged with lines, as well as the ball may ~~be~~ disappear which be hidden in the stadium or flying on air, and similar appearance of players, etc. The detect ball is the first step for tracking in broadcast soccer video. There are several methods of ball-tracking are based on their problem. In this paper, we have discussed different methods of object detection and tracking in the soccer videos which are available in the literature.

***Keywords*: **Object detecting, Soccer video, Object tracking, Kalman filter, Background subtraction, particle filter**,** Multiple hypothesis tracking.

1. Introduction

In computer vision, object detection is a very important step to identify the interesting objects in the image[1][2]. Over the past decades, although the ability to detect objects has improved greatly, it is still considered a complex problem to solve. There is a wide range of applications that depend on identifying objects such as sports video analyzing, surveillance, medical image processing, etc.[3] Object detection can be done by various techniques such as background subtraction, optical flow, and frame differencing[4]. This paper focuses on soccer sports videos.

Soccer videos get a lot of attention because of its immense popularity as it represents the most-watched and most-followed in the world with an estimated 250 million players worldwide.[5] Therefore, football analysts often want to analyze the game which used in wide applications, such as player action recognition, evaluation of the weaknesses or strengths of a player or a team, verification of referee decisions, statistical evaluations, goal analysis or pattern of attack, animations, visualization, 3D reconstruction of the soccer game, video compression, summarization, Creating and optimizing editorial material, Retrieving and indexing content, highlight detection, player trajectory extraction.[6]

The paper is organized as follows. In section 2 literature discussion about work related to object Detection and Tracking in Soccer Videos. In section 3 the tracking process is described. The conclusion of the paper is shown in section 5.

2. Related work

The soccer video contains much research related to ball detection and tracking; some of these research are described below:

Orazio in 2002 [7], proposed an algorithm for detecting a ball using background subtraction and modified Circle Hough Transform. This algorithm is modification by Tim approach [8], although using Modified the Atherton Algorithm for improvement the result, but his approach, still can't overcome occlusions and demands, therefore, its application still very limits. **Xinguo, et al in 2003** [9], presented a new method for tracking the ball in broadcast soccer video that includes two procedures, first one depending on the ball's attribute (shape, size, and color) to detect the ball, second depending on the Kalman filter to track the ball. About 89.85% was the highest achieved accuracy. **Xinguo, et al in 2004** [10] and **Dawei, et al in 2005** [11], presented the same schema for detecting and tracking ball offline in soccer video that contains two phases, one for generating a set of candidates balls and the other for computing the trajectories of the ball through using these candidates. For tracking the ball, **Xinguo** used only Kalman filter and get an accuracy of about 81%. While **Dawei** used template matching and Kalman filter which led to a high accuracy of about 89.7% and can be working well even in the bad playfield. **Joo, et al in 2007**[12], presented an improvement of MHT (Multiple hypothesis tracking) using a modification of Murty’s algorithm for tracking players By allowing one player to be identified with several measurements, and vice versa. As a consequence, the multiple hypotheses in the case of occlusions and noise outperformed the single hypothesis. The tracking mistake, however, occurred due to a sudden change in the player velocity in extreme occlusions or poor estimation of measurements within an occluding blob. **Yu in 2008** [13], presented a new approach to detect the ball and players. This approach includes three phases: extract playfield by using histogram learning technique to get rid of lighting and shaded areas, extract foreground blobs by morphological processing (erosion and dilation), eliminate false alarm (not ball, not player) by skeleton pruning and shape analysis. This approach was compared with two different approaches are [14] and [15]. The approach [14] content two steps: extract playfield by using color histogram, and Kalman filtering with template matching for tracking the players, while approach [15] is tracking the players using a color-based template matching. These approaches are failed for detecting the ball and finding players when they merge with field lines. The result was the proposed approach [13] is better to implement than the previous two approaches in detecting the ball and players. The proposed algorithm works perfectly only when players are far apart and not overlapping at the same place as well as the system able to track a small ball (as small as 30 pixels) and successful for handling the occlusion through using motion estimation and adaptive particle filter. The problem of this algorithm is if it detects the wrong ball that causes the tracking to fail. **Chiang et al. in 2009**[16], proposed a system for tracking soccer players based on mean shift algorithm and motion prediction. Discriminatory color selection is employed to model soccer players, and outgoing and incoming players are identified and managed in this system, but it could not handle occlusions among teammates or complete occlusions as well as it needed a part of the player within the initial search window. In the case of similarity background players colors or changes in appearance (such as changes in lighting), the system can fail. **Naushad, et al in 2012** [1], who proposed a new algorithm for detecting the players and ball which contains four steps: elimination of the ground by automatic ground detection algorithm to get rid of lighting and shaded areas, extracted the players and candidate balls by the Sobel gradient method, elimination of the line by line detection algorithm and elimination of the unwanted object by the threshold. This algorithm was compared with the algorithm of Jong-Yet al[17]. and the result of the proposed algorithm was stronger than the previous algorithm in detecting the ball and players. The proposed algorithm works perfectly only when players are far apart and not overlapping at the same place. **Sanyal, et al in 2016** [18], solved the problem of [13] by constructing a new algorithm based on adaptive particle filter comprising three phases: predicting the measurement area around a ball, weighing the predicted particles to calculate space points, and resampling the winner points depending on their weight from measuring space. Detection of the ball depending on edge, shape, and color of it. This algorithm was compared to the algorithm of [19] for human tracking. The result showed that the proposed was better than[19] approximately by 7.2% and enable to track the ball when there is partial occlusion. **Huda, et al in 2020**[20], presented a new system to detect the players and a ball (as small as 30 pixels) in real-time by using background subtraction and Sobel detection. Results have been more accurate and faster than using only background subtraction it approximate to two times faster with true detection which is approximate to %93. **Huda, et al in 2020**[21], proposed a new algorithm suitable for online to detect and track the ball. The proposed algorithm includes three important phases: 1) identify candidate balls instead of attempting to determine a ball in each frame to reduce missing rate balls and allow the overlap or merge balls as candidate balls. 2) determining the ball from the candidate balls through finding the true position by computing the distance between the candidate balls and ball which is dependent on the threshold, 3) an Extended Kalman filter is used to predicate, correct , and estimate the ball position. The proposed method was successfully implemented and the results showed that the algorithm can track the ball even if the ball's position is lost with high accuracy of approximately 92%.

3. Tracking Process

Object tracking is a significant action in video analysis. Video analysis can be done by three steps: detect the interesting moving objects, tracking these objects from frame to frame, and analysis objects tracking to recognize their behavior. [22]

3.1. Object Detection

In soccer videos, object detection is an important step to identify the interesting objects in the image which [23][24] can be done by various techniques, some of them are:

- **Background subtraction**

In this method, every image in a video sequence is compared with the background which is built by the median filter and Mean filter techniques used. By subtracting the current frame from the background, the pixel is considered as a part of the foreground if the difference between them is greater than a threshold T_s, otherwise considered as a part of the background[3].

$$B(x, y, t) = median\,\{F(x, y, t-i)\} \quad (1)$$

$$|F(x, y, t) - median\,\{F(x, y, t-i)\}\ | > T_s \quad (2)$$

Where, $i \in \{0, 1 \ldots n-1\}$

background subtraction classifies into two approaches: Recursive algorithm and Non-Recursive Algorithm.

In the Recursive algorithm, all the previous frames do not store into a buffer for background modeling, while in the Non-Recursive Algorithm all the previous frames stored into a buffer, therefore Recursive algorithm requires less storage than Non- Recursive Algorithm.

- **Frame differencing**

Moving objects is obtained by computing the difference between two consecutive images in video sequence[25]. The computation of this method is simple and easy for implementation by subtracting the current frame from the previous frame, the pixel is considered as a part of the foreground if the difference in the pixel values is greater than a threshold $T_{s.}$

$$|frame_i - frame_{i-1}| > T_s \quad (3)$$

3.2. Object Representation

The object representation is usually defined as consisting of a shape representation or appearance representation[26][27][28].

1. Shape Representation

It describes the shape of an object to make it is possible for detection and tracking. The commonly used shape representation includes points, geometric shapes, silhouettes, contour, articulated shape models, and skeletal models

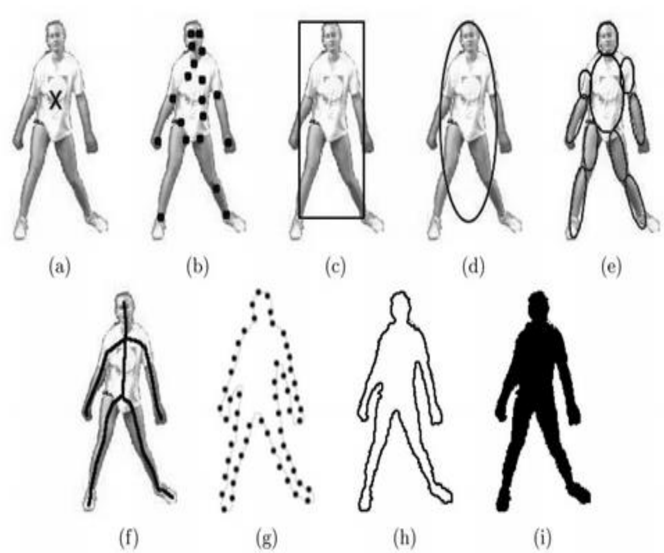

Fig. (1) : Various Ways For The Representation of The Shape

(a) Using a single point, (b) Using a set of points, (c and d) geometric shapes, (e) Articulated shape, (f) Skeletal model, (g,h) contour representation, and (i) silhouette representation [28]

- Points

The interesting object is represented either by one point or ~~a~~ several points. In a video, when tracking multiple objects by using several points may be a problem in the case of interaction between objects, e.g. full or partial occlusion. It is difficult to know the point belongs to which object for keeping the track, therefore, this way is suitable for simple, small objects which can be represented using a single point.

- Geometric shapes

Primitive geometric shapes are a common approach for representing simple shapes such as an ellipse, circle, square, or rectangle which is used for tracking vehicles and people. The disadvantage of geometric shapes is the parts of the background may be residing inside the defined shape or parts of the objects may be left outside of the defined shape.

- Silhouette and contour

It is a flexible model for representing non-rigid or complex objects, and many various object shapes, also called Blobs. It represents the region inside the contouring boundary or outline of an object.

- Articulated shape models

It is a way to create an object by grouping different parts together. By using simple geometric shapes like ellipses, can be represented every part. when representing, a human can be used this model by link hands, torso, arms, legs, head, and feet.

- Skeletal model

It is the most common model used in object recognition. This model extracts the object skeleton by Using the silhouette of an object and applying the medial axis transform.

2. Appearance Representation

There are many approaches for representing the appearance representation, also called descriptors, such as template and probability-densities of object appearance [27][28]. Where the most common way for probability-density estimates of the object appearance is a color histogram which has the main limitation that two objects with a very similar histogram may have a completely different appearance. Template suitable only for tracking objects whose shapes do not vary significantly during tracing ~~,~~ is formed using silhouettes or primitive geometric shapes.

3.3. Feature Selection

The most important task of object tracking is feature selection which must be unique to distinguish the target object from other objects, features, such as gradient, texture, color, etc., in object tracking, the feature must be extracted from the frame then compare with other objects to find the most similar object in the next frame[28].

3.4. Object Tracking

Object tracking has been a problem for many years, there are several major problems related to tracking such as difficult and rapid motions, change of appearance, illumination, scale, multiple objects tracking, and occlusion, therefore many tracker algorithms have appeared. In general, tracking systems were divided into three groups using the tracking approach: silhouette tracking, kernel tracking, and point tracking[22][24][27][28][29]. Each approach includes a lot of methods which are summarized in table (2.1).

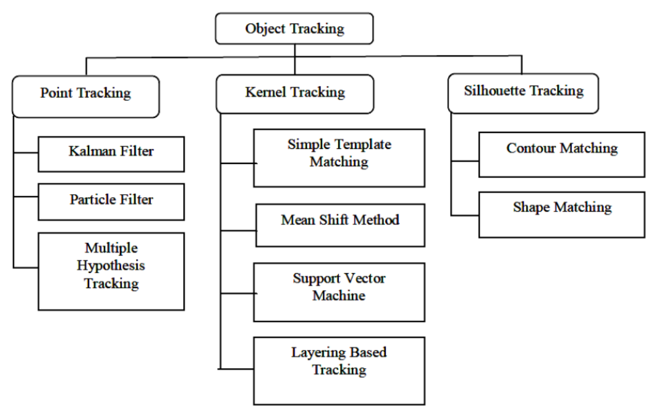

Fig.(Error! No text of specified style in document.**):** Types of Object Tracking [22]

3.4.1. Point Tracking Methods

Moving objects are detected and tracked from frame to frame based on their feature points. Large objects are represented by multiple points, then tracked in every frame. The disadvantage of this method is the complexity of tracking in case of there a lot of misdetections or points occlusions, therefore point tracking methods are used when detecting a small object to be represented by a single point [22][27][28][29][30].

There are two types of point tracking depend on the association of points are (1) Deterministic method (2) Statistical method. Deterministic meaning that no existing ~~any~~ randomness by giving an initial state or a start condition that always leads to the same output, while the statistical method uses the state space approach to model the object properties such as position, velocity, and acceleration therefore always contain noise.

Many algorithms depend on point tracking methods. Some of them are:

- **Kalman filter:** is used as an optimal solution for many tracking and predicting application which works optimally for linear models and Gaussian distribution. Kalman filter algorithm involves two stages, prediction and correction. Prediction is the first step that involves a prediction of the next state (velocity and position). The correction stage starts after the noisy measurement has been obtained. It incorporates an update of the Kalman filter that includes a state update and an update of the uncertainty (decreasing the uncertainty). If the function is a nonlinear function, then the initial state and noise become~~s~~ non-Gaussian distribution. To solve this problem, the function can be linearized using an extended Kalman filter.[31]
- **Particle filter:** is a recurrent algorithm-based computational method for tracking multiple objects using for approximating the posterior distribution by set weighted particles meaning this filter not required the system be distributed by a Gaussian. It performs two steps: prediction and correction as same as Kalman Filtering.
- **Multiple hypothesis tracking:** is an iterative algorithm based on a set of hypotheses. Each hypothesis uses to predict the next state, then actual measurements are compared with the predictions using evaluate a distance measure. This algorithm used for tracking multiple objects which can reach the optimal solution and handle the occlusion.

3.4.2. Kernel Tracking Method

Kernel Tracking is the most common method used based on object shape and appearance. It performs by computing the motion of the object from one frame to another, for determining its next position. For representing the object, Different primitive shapes such as an ellipse or rectangle templates are used. The disadvantage of this method is the parts of the background may be residing inside the defined shape or parts of the objects may be left outside of the defined shape[22][27][28].

Many algorithms depend on Kernel tracking methods which can be categorized into (1) Appearance Based Tracked (2) Multi-View Models. Some of them are:

- **Simple Template Matching:** is one of the most common methods in the appearance for finding small parts of an image in each frame which equivalents or match model with an image (template). Template represent an interesting region of the object which used for tracking using search algorithm such as brute force search. The disadvantage of this method is consuming time for complex templates. To overcome this problem, limiting the search to the nearest 8 or 4 neighborhoods.
- **MeanShift method:** is an iterative algorithm for finding the similarity by using Histograms. To track the object from one frame to next., first, the target object represented by elliptical or rectangular region, then finding the maximizes similarity score between the target and the current image region based on their feature that most commonly their color.
- **Support Vector Machine (SVM):** is mainly used for classification problems by giving a set of positive and negative training values. It gives a robust solution with noise. The positive values represent an interesting object which should be tracking while the negative values represent all the remaining things that should not be tracked[32].
- **Layering based tracking:** is another method of the kernel for tracking multiple objects which are based on intensity and motion in each layer such as layer appearance, rotation, translation. Layering is obtained by compensating the background motion to estimate all object's motion from the rewarded image, then the probability of every pixel be computed depend on shape features and object's foregoing motion. This method provides a robust solution with full occlusion of object[32].

3.4.3. Silhouette Tracking Method

Some objects have complex shape which cannot be represented by simple geometric shapes such as shoulders, fingers, hand. Silhouette method able to get an accurate shape description by finding the object region in each frame using object contour, object edge or color histograms[24] [27][28]. This method can be categorized into:

- **Contour Tracking:** Contour Tracking can be carried out using two different ways. The first way used state-space models contour motion and shape. The second way directly evolved the contour by minimizing the contour energy using direct minimization techniques such as gradient-descent. This method ~~is able to~~ can handle a large set of object shapes.
- **Shape Matching:** performance is similar to the template-based tracking in the kernel approach. This method used for tracking ~~the~~ only a single object by finding Shape Matching in two successive frames.

4. Conclusion

This paper presented a review of the various available tracking techniques. In a soccer game, building a robust tracker is not an easy task. as player tracking is often incorporated into other preprocessing steps, such as elimination of the shadow, player detection. There have been mentions of some effective tracker methods in the literature which have been built depend on many consideration, such as is it required for tracking ball or multiple players, or both of them? is it needed in real-time or offline? is it required to handle occlusions? is it needed to overcome the shadow and illumination problem, etc. Summarization of the tracking methodologies is illustrated in table1.

Table 1 : Summarization of the tracking methodologies.

(#: the number of objects tracking, where S: Single, M: Multiple, and symbols √ or × indicate whether the tracker requires or does not require training and does it optimal or not).

No.	Method	#	Category	Training Rule	Optimal
1	Kalman filter	S	Point tracking	-	√
2	Particle Filter	M	Point tracking	-	√
3	MHT	M	Point tracking	-	√
4	SVM	S	Kernel Tracking	√	-
5	Mean shift	S	Kernel Tracking	×	-
6	Template matching	S	Kernel Tracking	×	-
7	Layering based tracking	M	Kernel Tracking	×	-
8	Contour matching	S	Silhouette tracking	√	√
9	Shape matching	S	Silhouette tracking	×	-

Table 2 : Summarization of the Related Work

Year	Name of researcher	Method	Aim of algorithm	Result
2002	Orazio and others	background subtraction and modified Circle Hough Transform	Ball detection	very limits its application.
2003	Xinguo and others	Kalman filter	Ball tracking	89.85%

Year	Name of researcher	Method	Aim of algorithm	Result
2004	Xinguo and others	Kalman filter	Ball tracking	81%
2005	Dawei and others	template matching and Kalman filter	Ball tracking	89.7%
2007	Joo and others	Multiple hypothesis tracking	Players tracking	Tracking mistake, however, occurred due to a sudden change in the player velocity in extreme occlusions
2008	Huang and Joan	histogram learning technique and adaptive particle filter	Ball tracking	Work only when players are far apart and not overlapping at the same place. Limited uses because if the false ball is detected it leads to a failure to track.
2009	Chiang and others	mean shift algorithm	Players tracking	the system can fail when changes in appearance or similarity background players colors
2012	Naushad and others	Sobel gradient method	Ball and players detection	Work only when players are far apart and not overlapping at the same place
2016	Sanyal and others	adaptive particle filter	Ball tracking	better than Huang's schema about approximate 7.2%

Year	Name of researcher	Method	Aim of algorithm	Result
2020	Huda and others	background subtraction and Sobel detection	Ball and players detection	93.36%
2020	Huda and others	Euclidean distance between the candidate balls and ball, and Extended Kalman filter	Ball tracking	92.79%

References

[1] Ali, M.N., and Lee, S.L., 2012, An Efficient Algorithm for Detection of Soccer Ball and Players, Signal Processing Image Processing and Pattern Recognition conference, 39–46.

[2] Alya'a, R. A., and Dhannoon, B. N., 2019, Real Time Multi Face Blurring on Uncontrolled Environment based on Color Space algorithm, Iraqi J. Sci., 618–1626.

[3] Dhenuka, M., and Udesang, K., Hemant, D., 2018, Multiple Object Detection and Tracking: A Survey, Int. J. Res. Appl. Sci. Eng. Technol., vol. 6(2), 809–813.

[4] Christian, J., and Jensen, C., 2015, Event detection in soccer using spatio-temporal data, Master's Thesis, Comput. Sci., November.

[5] Dlgsci, Sports Dimensions Guide For Playing Areas, 6th Edition., June. 2016.

[6] Manafifard, H., Ebadi, M., and Abrishami, H., 2017, A survey on player tracking in soccer videos, Comput. Vis. Image Underst., vol. 159, 19–46,.

[7] Orazio, T., Ancona, N., Cicirelli, G., and Nitti, M.,2002 "A ball detection algorithm for real soccer image sequences, Proc. - Int. Conf. Pattern Recognit., vol. 16 (1), 210–213.

[8] Atherton, T. J., and Kerbyson, D. J., 2000, Size invariant circle detection, Image Vis. Comput., vol. 17 (11), 795–803.

[9] Yu, X., Xu, C., Tian, Q., and Leong, H., 2003, A ball tracking framework for broadcast soccer video, Proc. - IEEE Int. Conf. Multimed. Expo, vol. 2, II273–II276.

[10] Yu, X., Sim, C., Wang, J.m and Cheong, L., 2004, A trajectory-based ball detection and tracking algorithm in broadcast tennis video, Proc. - Int. Conf. Image Process. ICIP, vol. 5, 1164-1177.

[11] Dawei, G., and Yang, L.m2005, A Scheme for Ball Detection and Tracking in Broadcast Soccer Video, in Conference Paper in Lecture Notes in Computer Science, 865–875.

[12] Joo, S., and Chellappa, R., 2007, A multiple-hypothesis approach for multiobject visual tracking, IEEE Trans. Image Process., vol. 16 (11), 2849–2854.

[13] Huang, Y., and Llach, J., 2008, Tracking the small object through clutter with adaptive particle filter, ICALIP 2008 - 2008 Int. Conf. Audio, Lang. Image Process. Proc., 357–362.

[14] Seo, Y., Choi, S., Kim, H., and. Hong, K.,1997, Where are the ball and players? Soccer game analysis with color-based tracking and image mosaick, Lect. Notes

Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 1311, 196–202.

[15] Utsumi, H., Miura, O., Ide, K., Sakai, I., Tanaka, S., 2002, An object detection method for describing soccer games from video, in IEEE ICME.

[16] Chiang, T., Leou, J., and Lin, C.,2009, An improved mean shift algorithm based tracking system for soccer game analysis, in Proceedings: APSIPA ASC: Asia-Pacific Signal and Information Processing Association, Annual Summit and Conference, 380–385.

[17] Jong, Y. K., 2009, Soccer Ball Tracking using Dynamic Kalman Filter with Velocity Control, in Sixth International Conference on Computer Graphics, Imaging and Visualization.

[18] Sanyal, S., Kundu, A. and Mukherjee, D., 2016, On the (soccer) ball, ACM Int. Conf. Proceeding Ser.

[19] Jia, X., Lu, H. and Yang, M., 2012, Visual tracking via adaptive structural local sparse appearance model, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 1822–1829.

[20] Najeeb, H. D., and Ghani, R. F., 2020, Tracking Ball in Soccer Game Video using Extended Kalman Filter, in 2020 International Conference on Computer Science and Software Engineering (CSASE), 78–82.

[21] Huda, D. N., and Ghani, R. F., 2020, Shadow Elimination in Soccer Game Video Using Background Subtraction and Sobel Operators, Iraqi J. Sci., vol. 61(9), 2426-2433.

[22] Parmar, M., 2016, A Survey of Video Object Tracking Methods, vol. 4 (1), 519–524.

[23] Dutt, S., and Kumar, G., 2018, A Survey on Moving Object Detection and Tracking Based On Background Subtraction, Oxford J. Intell. Decis. Data Sci., vol. 1, 55–62.

[24] Shilpee, A., Dave, A., Nagmode, M., 2013, Statistical Survey on Object Detection and tracking Methodologies, Int. J. Sci. Eng. Res., vol. 4 (3), 1–8.

[25] Samhitha, P., Jyothi, A., Vesapogu, R., Mannem, M., and Harsha, S., 2017, Vehicle Detection , Tracking and Speed Measurement for Traffic Regulation, Int. Res. J. Eng. Technol., vol. 4 (4), 467–469.

[26] Verma, R., 2017, A Review of Object Detection and Tracking Methods, Int. J. Adv. Eng. Res. Dev., vol. 4 (10), 569–578.

[27] Ågren, S., 2017, Object tracking methods and their areas of application : A meta-analysis, Master Thesis Comput. Sci., 1-45.

[28] Yilmaz, A., Javed, O., and Shah, M., 2006, Object tracking: A survey, ACM Comput. Surv., vol. 38,(4), 1-45.

[29] Pathan, I., and Chauhan, C., 2015, A Survey on Moving Object Detection and Tracking Methods, vol. 6 (6), 5212–5215.

[30] Padmavathi, S., and Divya, S.,2014, Survey on Tracking Algorithms, Int. J. Eng. Res. Technol., vol. 3 (2), 830–834, 2014.

[31] Singh, H., 2018, Extended Kalman Filter: Why do we need an Extended Version?," lacture, [Online]. Available: https://towardsdatascience.com/extended-kalman-filter-43e52b16757d.

[32] Reddy, K., Priya, K., and Neelima, N., 2016, Object Detection and Tracking - A Survey, Proc. - 2015 Int. Conf. Comput. Intell. Commun. Networks, CICN, 418–421.