# Robust Visual Object Tracking


**Submitted by:**

Muhammad Faisal Bashir

01-241172-013

**Supervised by:**

Dr. Ahmad Ali

Department of Software Engineering

Faculty of Engineering Sciences

Bahria University Islamabad, Pakistan

# Robust Visual Object Tracking


Muhammad Faisal Bashir

01-241172-013

**Supervisor**

Dr. Ahmad Ali

A THESIS SUBMITTED TO THE DEPARTMENT OF SOFTWARE ENGINEERING, FACULTY OF ENGINEERING SCIENCES, BAHRIA UNIVERSITY ISLAMABAD IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER'S DEGREE IN SOFTWARE ENGINEERING

DATE: 31ST October, 2019

# Thesis Completion Certificate

Scholar's Name: Muhammad Faisal Bashir

Registration No.: 01-241172-013

Program of Study: Master of Science (Software Engineering)

Thesis Title:

Robust Visual Object Tracking

It is to certify that the above student's thesis has been completed to my satisfaction and, to my belief, its standard is appropriate for submission for evaluation. I have also conducted plagiarism test of this thesis using HEC prescribed software and found similarity index at _____ that is within the permissible limit set by HEC for the MS degree thesis. I have also found the thesis in a format recognized by the Bahria University for the MS thesis.

**Co-Supervisor's Name:** Dr. Raja M. Suleman

**Principal Supervisor's Signature:** _____

**Date:** _____ **Name:** _____

# Plagiarism Undertaking

I, **Muhammad Faisal Bashir**, solemnly declare that research work presented in the thesis titled **"Robust Visual Object Tracking"** is solely my research work with no significant contribution from any other person. Small contribution / help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and Bahria University towards plagiarism. Therefore, I as an Author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred / cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS degree, the university reserves the right to withdraw/revoke my MS degree and that HEC and the University has the right to publish my name on the HEC / University website on which names of students are placed who submitted plagiarized thesis.

Student/Author's Sign: _____

Name of the Student: Muhammad Faisal Bashir

# Author's Declaration

I, **Muhammad Faisal Bashir**, hereby state that my Master's thesis titled **"Robust Visual Object Tracking"** is my own work and it has not been previously submitted by me for taking any degree at this University **Bahira University Islamabad** or anywhere else in the country/world. At any time if my statement is found to be incorrect even after my Graduate the university has the right to withdraw/cancel my MS degree.

_____

Muhammad Faisal Bashir

Date: _____

# Abstract

*It is very easy for humans to recognize the object and to follow it till their sight range, but with the advancement of technology, we want to take this work through machines so that we can get better results according to our desire. So, with this need, the computer vision field came out of the box by using its sub-fields like object detection and visual object tracking. Tremendous efforts are being done by researchers in a field of object tracking, but it is still open to be explored because the challenges of visual object tracking still exist and this thesis also deals with the visual object tracking challenges. We consider the main challenge of tracking by our motivation i.e., Occlusion and illumination variation. So, in this regard, we select the state of the art algorithm name with "Adaptive Correlation Filters with Long-Term and Short-Term Memory for Object Tracking" in which three correlation filters were proposed 1) Long-Term Filter 2) Translation Filter 3) Scale filter. These filters work outstanding in most of the video sequence, but we have found that their performance degrades for some of the video sequences bearing challenges of occlusion and illumination variation. In order to solve these problems, we incorporate one more filter that is Kalman filter to the algorithm; enhanced algorithm yields better results as compared to its counterpart method when video sequences having challenges of occlusion and illumination variation is given to the proposed method has been tested on standard dataset i.e., Object Tracking Benchmark 13 containing 49 video sequences with different challenges the comparison of the proposed method with its base method i.e., existing selected method. The proposed method highlights its effectiveness both quantitatively and qualitatively, especially in occluded and varying illumination environment.*

# Dedication

*This thesis is dedicated to my parents for their love, endless support, and encouragement*

# Acknowledgments

First and foremost, I am very grateful to Allah Almighty for helping me in every aspect of my life and in this thesis as well, paving my way to success.

Special thanks to my thesis supervisor **Dr. Ahmad Ali** for his support, motivation and valuable time. The door to his office was always open for me whenever I had trouble and needed his advice. I would also want to acknowledge my parents, especially my mother who always prayed for my success. Furthermore, I would also like to acknowledge the assistance of friends and mentors from the **Aviation Design Institute, PAC Kamra**.

# Table of Contents

# LIST OF FIGURES

# List of Tables

# List of Abbreviations

| | |
|---|---|
| VOT | Visual Object Tracking |
| AR | Activity Recognition |
| CV | Computer Vision |
| CFT | Correlation Filter Trackers |
| NCFT | Non-Correlation Filter Trackers |
| ROI | Region Of Interest |
| CNN | Convolutional Neural Network |
| 3GSS | Third generation surveillance system |
| KCF | Kernelized correlation filters |
| R-CFTs | Regularized Correlation Filter Trackers |
| PLF | Pixel-Level Fusion |
| MDNet | Multi-domain Network |
| ECT | Enhanced CNN tracker |
| GOTURN | Generic Object Tracking Using Regression Network |
| RDM | Reinforced Decision Making |
| PMGRT | Part-Based Multi-Graph Ranking Tracker |
| CEST | Context-aware Exclusive Sparse Tracker |
| FCN | Fully Convolutional Network |
| TACF | Target-Aware Correlation Filters |
| HOG | Histogram of oriented Gradients |
| MCPF | Multi-Task correlation particle filter |
| OTB | Object Tracking Benchmark |
| MOSSE | Minimum Output Sum of Squared Error |
| K.F | Kalman Filter |
| $A_T$ | Translation Filter |
| $A_L$ | Long-Term Filter |
| $A_S$ | Scale Filter |
| FPS | Frames per Second |

## Chapter 1

# Introduction

For human it is very easy to understand the object and detect the desired target or to pursue that object continuously, but with the evolution of technology this becomes the need to manage this whole work with some sensors (e.g. Cameras) so the computer vision field is introduced to deal with this type of work, which deals with the automatic extraction, understanding and gathering the useful information from a single or multiple images. Computer vision has various fields, but our research work is related to the tracking of objects which is a subfield of CV, the main objective of VOT is that to find the specific point in an image which is known as RoI (Region of Interest) by providing some coordinates as input. The usability of VOT in real-world systems is very wide [1] few applications are shown in Figure 1.1.

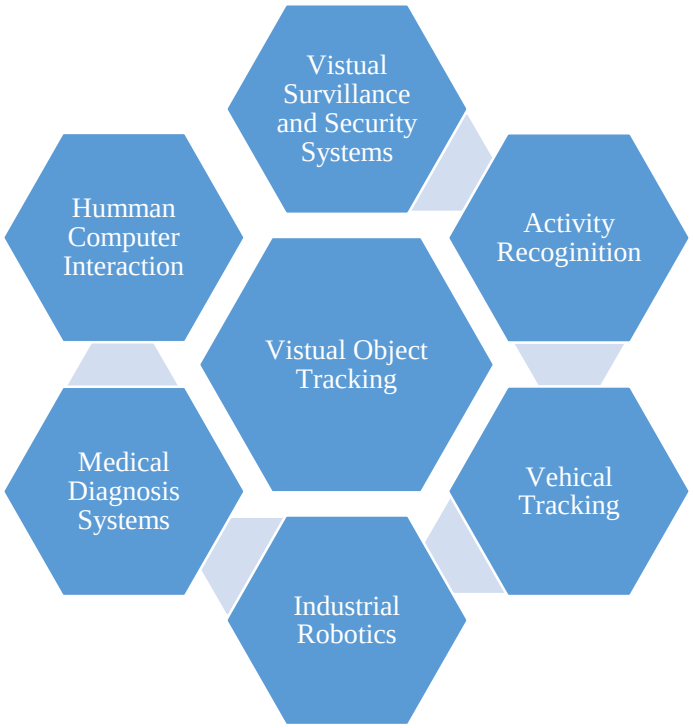



**Figure 1.1:** VOT Application

- ***Visual Surveillance and Security Systems:*** VOT is a crucial part of clever visible surveillance. These systems are globally used, surveillance of different locations and homes associated with public and defense sectors [1] for unusual activities and their

detection e.g. 3GSS-Third generation surveillance system [2], CX EDS-Siemens Sistore [3].

- ***Activity Recognition:*** VOT can be applied for activity recognition in internal and external environment e.g. Human Activity recognition [4]
- ***Vehicle Tracking:*** VOT is also applied to vehicle tracking e.g. the tracking of vehicles by unmanned aerial vehicle [5]**,** the autopilot of a UGV [6].
- ***Industrial Robotics:*** Control systems consist of human-made robots and industrial robots are used for VOT e.g. humanoid robot ASIMO [7] and visual control for UAVs[8].
- ***Medical Diagnosis***: As VOT has emerged in every field so its roots are growing in the medical field as well for the identification of multiple diseases e.g. ventricular wall tracking [9] and vocal tract shape reconstruction [10].
- ***Human-Computer Interaction:*** To make a community lifestyle better and make ease of use to interact with machine VOT emerged as a vital role in the community e.g. sixth-sense (a wearable gesture interface) [11] and eye gaze tracking for disabled people [12].

VOT attains much reputation due to its extensive use in different applications. In the process of VOT, one must find the region of interest in a sequence. Object annotation is done in the target initialization process through different symbols: bounding box on an object, ellipse, centroid, object Skelton or object silhouette. The target position is determined by the boundary in a rectangular shape, given as input in the first frame[13]. Literature shows that object tracking is an interesting topic and various surveys have been published.

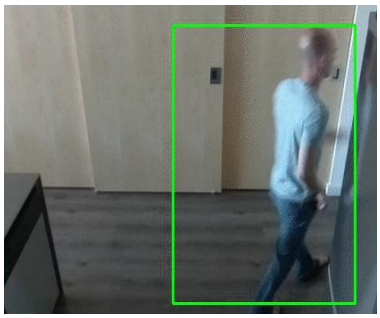
## 1.1. Visual Object Tracking Challenges

VOT application is frequently used, as they have many advantages, but besides this, there are also numerous challenges shown in Figure 1.2 while doing VOT [1]. Some issues are briefly explained below

**(a) Occlusion**

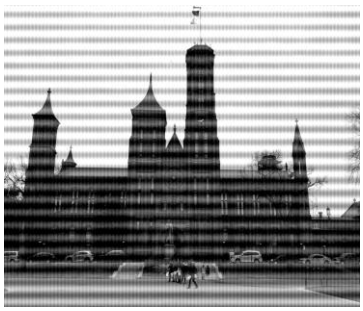
**(b) Changing Appearance**

**(c) Cluttered background**



**(d) Changing target size in the image**

**(e) Illumination variations**



**(f) Noise in the image**

**(g) Similar objects**

**(h) Complex object motion**

**Figure 1.2:** VOT Challenges

(a) **Occlusion:** It is the condition of the object where the object entirely hides by something, we may be partially hidden by something. There is no such technique to resolve this issue only best practices can be adapted according to the scenario of the environment.

(b) **Changing Appearance:** The selected target most of the time changes their appearance during motion, appearance is changing during motion, so it is very difficult to maintain the model with that appearance, to update model frequently is a challenge.

(c) **Cluttered background:** This is the most severe problem in VOT if the targeted object has too many objects in the background then it is known as a

cluttered background. There are two environment types in which target is identified, one is indoor and the other is outdoors. Tracking in an outdoor environment is difficult as compared to an indoor environment where we know about the details of the environment.

(d) **Change of target size:** During tracking, there is a possibility that due to movement of one thing, Object changes their position some time camera move towards the target the size of target increases and if moves away from the target then the size of the target is decreased.

(e) **Illumination variations:** Illumination is an important aspect of tracking. But besides this, this is the main headache because some necessary feature which our model knows is prominent in high illumination and, but they are not prominent in low illumination and vice versa. So, we must consider the illumination as an important factor in the whole process of tracking.

(f) **Noise in the image:** Sometimes it may happen that due to faulty sensors we get a noisy image in which some important information (e.g. color) is missing.

(g) **Similar objects:** It is very difficult to track the target if there is a similar object exist in that environment.

(h) **Complex object motion:** It is very difficult to track that target which is continuously changing its position, moving from one point to another haphazardly and quickly. May be target goes out- of-plan because it changes its direction in too much speed (e.g. Fighter jet motion).

## 1.2. Motivation

As war and terror from the last couple of years were increased and security surveillance systems are at stake, so it is very necessary to bring improvement in security surveillance systems. So this is the main motivation as equally there is advancement in technology take place across the earth, human wants to stimulate each, and everything done on their fingertips this motivates more than one field to introduce, from one of these field computer vision has emerged as a big monster with its subfield visual object tracking, so automatic video/image detection of some specific object becomes more interesting and powerful full application of this subject. After studying the literature, we came to know that to track an object in a specific video sequence or through any

surveillance camera is really a big challenge to accomplish. As there is a huge number of data present in a video so it's a very time-consuming process.

Tracking gives an opportunity to extract useful information like object detection, object classification, human identification and scene motion, etc. But besides these functionalities, there are multiple challenges that become a hurdle to get a good result in tracking, therefore, many types of research are motivated to do work in this open and wide research area.

## 1.3. Problem statement

From previous sections, it is found that VOT is an interesting field for researchers as well as it's very important, but there is a lot of issues still exiting in VOT as explained VOT challenges section so still, it is an open research area for researchers. Most of the occurring problems are unknown target, track multiple targets, track target by multi-sensors, tracking in static or dynamic environment, online and offline targets, so our objective is to propose novel heuristics to support valuable tracking by good criterion. We also aim to build on previous work to propose a strategy to generate a better tracking solution. For the solution, we also analyze the difference between proposed and existing strategy or brings improvement. Finally, we conduct an empirical evaluation to check the performance of existing tracking algorithms in solving multiple constraints and to generate a solution corresponding criterion. Our main target is to address the occlusion and illumination variation (high or low both) challenges.

## 1.4. Research Methodology

In the whole thesis, stepwise research methodology is followed

1. **Literature Survey:** The main purpose to analyze the state of the art techniques by performing the literature survey. We get familiar with the various approaches of tracking exist in literature.
2. **Formation of Novel Heuristics:** In a second step after analyzing the literature review, we move towards our main goal that was to provide an effective and efficient solution
3. **Empirical Evolution:** Finally, we evaluate the heuristics on the existing dataset to check the performance of the existing (the selected one) and the improved solution provided by us.

## 1.5. Contributions

This thesis brings out the following contributions

1. Improve the tracking performance of long and short term tacking by using the recently proposed technique with some conventional technique.
2. Provide an efficient solution to detect objects facing occlusion and illumination variation challenges of VOT.
3. To analyze some object tracking methods for the single object.

## 1.6. Document Structure

The remaining thesis is ordered as that:

In Chapter 2 we do a detailed literature review to understand the VOT and its problems to explore the limitation in existing work so that we can bring improvement to that work.

In Chapter 3 a proposed tracker is presented that what improvements we bring to the existing tracker and discuss all the existing and proposed a technique in detail.

Experiments are performed in Chapter 4 in which quantitative and qualitative experimental results on the existing tracker and proposed tracker are discussed.

Chapter 5 discusses the conclusion of all the work we have done with the thesis and future work is proposed.

# Chapter 2
# Literature Review

In this chapter literature review and its analysis are provided. All available studies which are closely related to our research is included. Firstly, we classify the related literature. Secondly, we present concepts and methods related to tracking including object detection, object detector, convolutional neural network, and MOSSE. After that, we explain the analysis of related studies is presented. In the end, we conclude this chapter by selecting the dataset for experiments.

## 2.1. Classification of Tracking Approaches

Tracking methods are classified into two categories:

1) Traditional Approaches
2) Modern Day Approaches

Visual object tracking is an interesting field of research a lot of efforts done by the researcher for the last four decades. Researchers are working in this field by different approaches, some are traditional, and some known as modern-day approaches [1]. Traditional approaches are very important approaches because they provide the basis for the modern-day-approaches. Few approaches are briefly explained to understand the basic phenomena of visual object tracking in a better way. However, all types of approaches are shown in Figure 2.1.

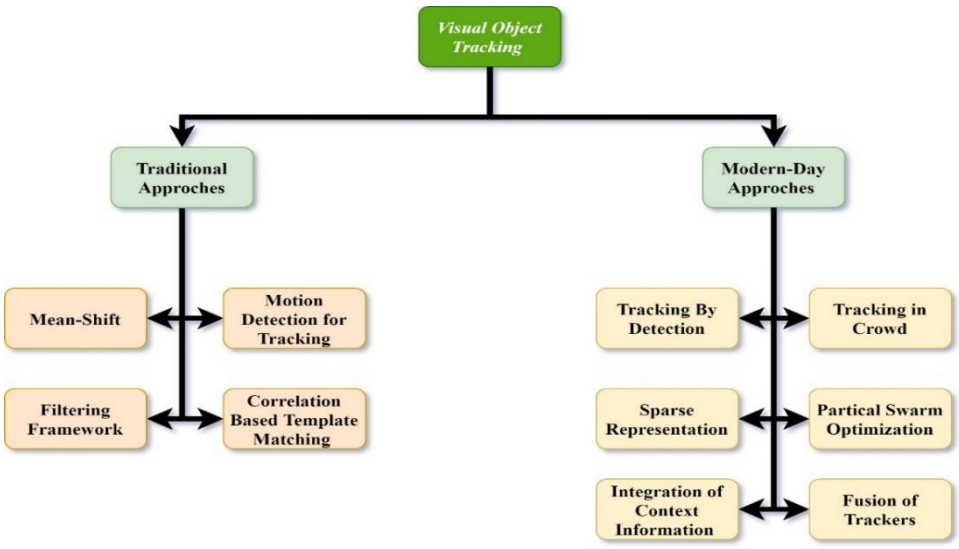



**Figure 2.1:** Different Object Tracking approaches

### 2.1.1. Traditional Approaches

#### *2.1.1.1. Mean Shift for VOT*

It stands as an iterative technique used to discover a mode of distribution provided. It is a simple and straightforward algorithm. Some several images given to this randomly as an input and image pixel as representativeness of cluster focus. Mean shift-based schemes face some problems as they required manual modifications of system parameters.

#### *2.1.1.2. Filtering framework*

Another approach used for tracking of objects is filtering. A recursive numerical parametric procedure is intended just for the distinct time systems. KF is mostly used with other algorithms in conjunction. KF works usually in two modes

1. ***Normal tracking mode:*** KF is used to predict the next location of targets in the given video sequence.
2. ***Occlusion mode:*** In this mode, KF use predicted value and ignores the measured value.

Mostly researcher uses KF for the discriminative tracking approach. KF is used for the assumption of linear systems and when Gaussian measurement is measure than the noise is true always.

#### *2.1.1.3. Correlational Based Template Matching*

This technique is identified at the beginning of the computer vision field. In the process of tracking the target is identified in the first frame manually or automatically. Target is represented by a template, which is used for correlating in the video frame. The new target position is identified where the correlation score is high. Updating the template is necessary because the target changes its position very rapidly.

#### *2.1.1.4. Detection of Motion for Tracking*

Here are the following techniques for the detection of motion [1].

1. *Background subtraction*
2. *Temporal Difference*
3. *Optical flow*

1. **Background Subtraction:** Anything else in the scene except the Target of Interest is named as background and the target of interest is known as foreground. Literature shows that background subtraction is used for two purposes, initialization of target and for the detection of target of interest. In the case of a fixed background, a frame is subtracted for detection. This method fails in outdoor environments where the background changes repeatedly.

2. **Temporal differencing:** To detect movement of objects in a scene in which the previous frame is subtracted from the current frame is known as temporal differencing. Target is not predicted as a foreground object when the target stops moving.

3. **Optical Flow:** In segmentation and tracking applications, optical flow is considered as the main feature. In optical flow, the motion which is a relative to both object and sensors used for capturing the object is being observed, with the help of consistency in brightness calculation is done.

### 2.1.2. Modern-Day Approaches

Traditional object tracking methods still have limitations [1] so most of the work in modern-day approaches is done on trackers in latest research so we see that in recent studies tracking algorithms are separated into dual primary groups: Correlation Filter Trackers (CFT) and Non-CFT (NCFT) with several subgroups in individual class [13].CF is used for the robustness and efficiency of tracking. DCF is mainly used in CV applications. In CF algorithms follows "tracking by detection" architecture.

**A) Correlation Filter Tracker**

*1.* ***Basic correlation filter-based trackers***: used kernelized correlation filters (KCF). Many other tracking algorithms use KCF as their base. Usually, the tracker uses other features such as HOG, color names or deep features (RNN, CNN) [14]. A Gaussian kernel function is used for the distinguished target objects and their surroundings in the KCF [14] algorithm.

*2.* ***Regularized Correlation Filter Trackers:*** Response maps in DCF have a precise score near to the center and additional scores are prejudiced of episodic guess causes degrading in performance. Regular-shaped target objects are learned by DCF. Due to the equal size of the patch and the filter DCF algorithm has a limited range in detection. Thus, it is necessary to include a measure for limitation of regularization and

this type of tracker remain as **Regularized Correlation Filter Trackers (R-CFTs)** [14].

*3.* ***Siamese-Based Correlation Filter Trackers:*** CNN is used to find similar features between two video sequences with the help of convolutional layers. This architecture is usually observed by the Siamese. The main objective of Siamese-Based CFTs is to find the weather identical object exits, or not, in two images given as input to the network, with the help of this network two inputs are joined to produce one output[13]. The siamese concept was initially used for different purposes of signature and fingerprint identification but later it is used for face identification [15][4]. Siamese are used to handle tracking challenges and incorporation of CFTs with Siamese network for VOT is classified as Siamese-Based CFTs[13].

*4.* ***Part-Based Correlation Filter Trackers:*** As the name shows part-based means those filters which learn target appearance in parts. Due to intra-class variability, different challenges of VOT may appear in a sequence[13]. In occlusion, a tracked object is may be occluded by some of another object. Part-based techniques are used in numerous applications containing detection of objects, Detection of pedestrian and recognition of human faces. Many part-based trackers are proposed and developed by the researchers to solve the problem of occlusion challenges.

*5.* ***Fusion-based Correlation Filters Trackers:*** Fusion means a combination, necessary facts are used to show progress in the performance of different applications of tracking. At a three-level image, fusion is achieved as Pixel-Level Fusion (PLF), Feature-Level Fusion and Decision-Level Fusion. Infrared and visual images are used for PLF by researchers and titles of colors are used as a feature for FLF [13]. Nonlinear regression kernels and tracking by detection is performed by KCF, on the given data (training and testing) data on KCF exploits on low computational cost.

**B) Non- Correlation Filter Tracker (NCFT)**

Those trackers which do not use correlation filters are considered as NCFT and we categorize Non-Correlation Filter based Trackers (NCFTs) into the following types:

*1.* ***Patch Learning Trackers:*** In general, a target is trained on both negative and positive samples. These trackers have achieved in both trackers and background patches. In Multi-domain Network (MDNet) the quantity of convolutional layer and fully connected layers are three and two respectively [16]. In the duration of on-line monitoring, primary frame layers are independently discovered. A local structural

information and inner geometry are exploited by the used of particle filter framework imposed by the Convolutional Networks without Training (CNT) proposed by Zhang et al [17] CNT is adaptive algorithm an effective representation is done due to its categorized manner because it has only two layers of feedforward in convolutional network. Short and long-term detectors contained by Exemplar based Linear Discriminant Analysis (ELDA)[18] tracker. The performance of ELDA is improved by a combination of Convolutional NN it is named Enhanced CNN tracker (ECT).

*2.* ***During the Multiple Instance Learning-Based Trackers:*** Object detection is done by MIL and it is extensively used in computer vision applications. The concept of bags is used, during the process of training samples are placed in bags rather than considering them as individual patches and levels of bags are decided as positive or negative through labels. A negative bag has all the negative samples, but if there is only one positive-sample in a bag than it is known as a positive sample. Instances of labels are unknown, but bag labels are known. Week and strong classifiers are made on the basis of instances. In paper [19] MIL framework is proposed using fisher information with tracker to choose weaker classifiers. Instead of log-likelihood, fisher information creation is used for unlabeled data, with the help of fractal dimensions and position distance significance of instance is calculated. A chaotic map is used for maximizing the likelihood of bags. An image is converted into a vector form and by zero-mean it is normalized the chaotic information is encoded [19].

In a paper [20] approach Mahalanobis distance is used for the computation of instance significance and with gradient boosting classifiers are trained. The importance of instances and bags is defined by Mahalanobis distance. The difference between positive and negative bags and maximum margin between the decided the week classifiers [20].

3. ***Siamese Network-Based NCFT Trackers:*** It is working on matching mechanism. Templates are matched with candidate samples to produce the likenesses between the patches. In paper [21] a new tracking technique is proposed, with name Generic Object Tracking Using Regression Network (GOTURN) in which during the process of tracking search regions and templates are cropped and after the cropping is done, they have context information, these two are fed to five individual convolutional layers, this tracker is offline forward feed which doesn't need directly regresses target location.

A new model Reinforced Decision Making (RDM) is made up of network policy and matching, with the help of matching network heat maps are generated and network policy is accountable for that production of regularized scores for the forecast of heat maps. Prediction maps are produced with the fusion of deep features, these deep features are obtained when the cropped search patch and the number of templates are forwarded to two separate convolutional layers. The decision about dependable Reinforcement learning is attained with the help of two layers of convolutional and FC layers contained in the policy network. Due to the prediction map we able to see the target maximum scores.

*4.* ***Superpixel Based Tracker:*** A group representation of the same pixel values. For the discrimination purpose classification is done on a superpixel beside this ROI is also segmented on superpixel. A lot of applications have been developed with the help of superpixel it has a lot of attention in the computer vision community [13]. In a paper [22] at a coarse-level and fine-level superpixel, Bayesian tracking method is given in which a confidence measure defines whether or not that superpixel corresponds to background/target and a few superpixel is computed by coarse-level appearance model so that only single superpixel is available in a given bounding box. More superpixels in the target area created on the target position in the earlier frame. The structural information with superpixel and preservers the basic target properties are explained by Structural Superpixel Descriptor (SSD) [23] in the SSD process greater weight is assigned to the superpixel which is near to target center and the decomposition of a target into different size of superpixel.

*5.* ***Graph-Based NCFT Trackers:*** As we know that the graph is made up of edges and vertices unlabeled vertices in the graph show the prediction labels. Applications of procedures made up of graphs are secondhand in many ways one of its best use in object detection [24]. Object appearance is represented by superpixel as a node and the inner geometric structure is represented by edges. A Geometric hypergraph Tracker (GGT) [25] is used to construct geometric hypergraphs by using the higher-order geometric relationship between target parts with correspondence of hypothesis a relationship between target and candidate part is represented. When the target and candidate parts are matched reliable parts are computed from correspondence hypotheses.

*6.* ***Part-Based NCFT Trackers:*** To handle deformable parts part-based modeling actively used in Non-CFTs. In paper [26] a part based tracker is proposed in which

object objects are decomposed into parts and every part has its own adaptive weight. Vertices represent parts and edges define consistent connection in a spanning tree which is further used for structural constraint. Distinguish between target and it's apart from the background is done by using online structured learning using SVM, the new target position is identified by the score of parts and target and this score is known as classification maximum score. Relationship among parts is used and target appearance is described by local covariance. Target is further separated into non-overlying parts. The illustration of a target, max pooling is used by a combination of multiple local covariance. Star graph is used for modeling of parts and the center part of the target is representing a central node. For the solution of linear programming problems during the tracking candidate parts pools and template, parts are used. The weighted voting mechanism is used for the selected part of the target, it is based on the relationship between the center part and surrounding parts [13]. In paper [27] presents the construction of graphs to rank target parts Part-Based Multi-Graph Ranking Tracker (PMGRT) is used. At the time of tracking, a target is broken in two things one is features and the other is parts. On the basis of parts and feature extracted multiple graphs are constructed, A weight matrix is constructed in which a graph is represented as features in rows and parts are represented in column.

7. ***Sparsity Based Trackers:*** We have studied discriminative tracking methods till now, but there is also a generative method as well one of its examples is sparse representation it has many applications in CV and image/signal processing. Minimization of reconstruction error and enough sparse is the main objective in this algorithm. During the tracking, previous knowledge is used to accomplish a difference between background patches and targets. In paper [28] Local and global target patches are jointly used to learn the sparse representation, and this all is done with the help of structural sparse tracking (SST) which is based on the particle filter framework. The high resemblance score is obtained through all of the particles and the target from the target dictionary is estimated by SST. The limitation of SST is that changed target patterns are selected by local patches due to occlusion or noise. In paper [29] context information utilizing the particle filter framework exploits the Context-aware Exclusive Sparse Tracker (CEST), and the representation of particle is done in a linear combination. With the help of the target template dictionary, a new target is estimated as best. The dictionary is modeled as a group to resolve the issues of tracking. In paper

[30] for the integration of discriminative and generative models, Hierarchical Sparse Tracker is proposed. It contains three types of histograms, having different functionalities. Local Histogram Model (LHM) encodes spatial information of targets, Sparsity based Discriminant Model (SDM) is responsible for the computation of target template sparse representation, and Weighted Alignment Pooling (WAP) is used to assign weights based on the similar features of target and candidates.

In this paper [31] robust and new tracking method is proposed. This method uses a fully convolutional network (FCN) so that the path is optimized by using dynamic programming and object probability map. Object appearance variation is solved, and occlusion is the deal with DP. Fixed feature extraction techniques are used in traditional approaches when the object changes it difficult to track the object through the fixed algorithm is composed of convolutional layers and recently used in segmentation and classification is a fundamental optimization technique and commonly used for tracking. Convolutional Neural Networks are made up of three components: 1) Convolutional Layers 2) pooling layers 3) Fully connected layers. For the learning of multiscale features upsampling is used with skip connections to learn. Feature maps are obtained by when filters are applied from input to the layer, with the help of backpropagation filters are learned and weighted. When feature maps extracted and fed into a pooling layer to lessen the computational time for future layers.

A target template and a search region are formulated in a convolutional feature. An accuracy gap is still being considered in and this gap is compared with extraordinary procedures such as ResNet-50 or deeper, with the help of analysis and experimental validation restriction is broken with a yet effective spatial aware sampling strategy and ResNet is trained successfully. A new architectural model is suggested in which accuracy is improved and the model size is trimmed [32]. The best answers are shown on the tracking benchmark and VOT2018 is one of them. The Siamese trackers are employed for the VOT problems in which general similarity maps by cross-correlation between the feature representation learned for the target template and searching for a neighborhood. In this paper, it is observed that all the networks build upon similar to Alex Net [33]. These types of trackers are used for the correlation problem and provides better results for deep networks. Execution will be further increased if it is handled by deeper networks. Training of deep networks is shown in on Siamese tracker is shown in this story.

The best object tracking frameworks are nowadays are correlational filter [34] on the basis of these CF many effective trackers being established spatially regularized CF tracking (SDRFC), being made for the solution of borderline belongings of CF tracking to improve performance of tracking used in SRDCF as regularization map to handle deformation or occlusion. A new scheme is proposed as energetic saliency-aware regularized correlation filter tracking (DSAR-CF) with the help of efficient, level-set algorithm available informing of the regularization weight map is done, which improves the accuracy and speed, with the help of the regularization weight map an appearance of the target is highlighted. The integration of the object dissimilarity information into the spatial weight map is the main task of this idea proposed in paper to boost the performance of the regularized Correlation Filter. On cropped region saliency detection with existing algorithm [35] is performed to overwhelm the background clutters and distractors. Experiments are conducted on a standard benchmark of object tracking OTB-13, OTB-15, and OTB-16. No effect is noted in the tracking speed of SRDCF. Improvement is done in an optimization method by replacing the gradient descent method.

The standard RGB-D trackers treat objects as 2D structure, due to which model work even in two out-of-the-plan rotation challenges. This limitation is handled by the OTR-Object tracking by reconstruction, due to robust online 3D target reconstruction is done by the set of view-specific discriminative correlation filters(DCFs). Recently proposed constrained correlation filer CSR-DCF [37] is a good contribution to the research community but OTR performs outclass due to enhancement of two following features

1. Due to heavy occlusion storage of view, specific DCFs which robust the target localization and point-cloud based estimation of change for 3D selection.
2. For 2D projection, accurate spatial support for constrained DCF is generated.

DCFs are efficient in both target localization and target appearance model. By considering the limitation of RGB and RGB-D trackers regardless of method (e.g. DCF[38], Siamese deep nets [39]) that the treat the target object as 2D. So the main contribution in this work is that OTR – Object Tracking by Reconstruction is proposed which reconstructs the 3D model with the view-specific DCFs attached.

In tracking community, Discriminative Correlation Filter (DCF) gets too much popularity because the targeted object is aligned with the targeted boundary in a rectangular shape. Due to an anomaly of shape Learned CF is automatically worsened

by the background pixel inside the targeted boundary to resolve this difficulty a Target-Aware Correlation Filters (TACF) for visual tracking is proposed [40] in which a target likelihood map is presented to impose discriminant weights. Further, a new optimization approach is proposed grounded on the preconditioned Conjugate Gradient method for effective filter learning with a hand help of handcrafted features Histogram of oriented gradients (HOG) with 31-dimensional features with 4x4 cell size. A MOSSE tracker [38] accomplishes an impressive tracking speed later different variations of CF are proposed to improve the performance of multi-dimensional features.

In paper [41] for robust visual object tracking a ROI pooled correlation filter (RPCF) is proposed in which through mathematical derivation ROI-based pooling can be equivalently achieved by enforcing the additional constraint on learned filter weights.ROI method performs pooling operations on the cropped ROI regions. An efficient joint training formula for the proposed correlation filter algorithm and drive the Fourier solvers for efficient model training, for the understanding of algorithm primal correlation is introduced first. This tracker is also evaluated on OTB-13, OTB-15 and OTB-17 benchmark with the one-pass evolution as evolution metric. The polling operation has been used in various fields of computer vision e.g. feature extraction [42, 43] and convolutional neural networks [44, 45]. An alternative solution for the ROI-based pooling with the circular constructed virtual samples.

In this paper [46] multi-task correlation particle filter (MCPF) is offered for the durable VOT. In MCF the different parts and features are jointly used to learn the CF and strength of MCF, and particle filter is represented. In [47] an online improving tracking method to update features to excuse for large appearance. Usually, discriminative trackers are using the classification technique to distinguish the target from its background [48].MCPF is evaluated on OTB-13 and OTB-15 datasets and its effectiveness is outstanding with the different revolutionary trackers. The four merits available in proposed MCPF are as follows:

1. Exploit the interdependencies among different features and combination of them to enchase each other responses
2. It handles the partial occlusion via a part-based representation.
3. For object detection, estimation different scales are drawn via sampling scheme and it handles large scale variation

4. MCPF spearheads the sampled particles toward the mode of the target state distribution and effectively covers the object sate well.

In this paper author purposed adaptive correlation filters by the combination of short and long-term memory for VOT. Attribute for effective VOT has three characteristics**. First,** the Correlation filter algorithm can achieve high tracking speed by calculating the spatial correlation efficiency in the Fourier domain. **Second,** CF naturally takes the nearby visual context into explanation and provide more discriminating data. **Third,** knowledge of CF is equal to a regression problem, where the circularly shifted version of image patches are regressed to soft labels e.g. Gaussian function with a narrow bandwidth ranging from 0-1. Contribution of this work

1. ACF is capable for the estimation of scale and translation changes and determine the failure occurs.
2. The effect of different feature types and the size of context are for designing the effective CF and ablation studies to the contribution of design choice.
3. Discuss, compare and evaluate the offered algorithm with the parallel work and revolutionary trackers on both data set OTB 2013 and OTB2015.

As correlation filter-based trackers [49] attain the extraordinary performance in benchmark evaluation and for the efficient utilization of visual tracking correlation filters are used [50]. Features can be extracted with the help of HOG an HOI, the offered HOI features bear some likeness to the distribution field scheme [51, 52] in which the characteristics of numerical values are shown as features. As the proposed method uses deep features, but this tracker is failed to track in high illumination. This tracker is evaluated on OTB-2013 and OTB2015 it performs well in some cases but not be able to perform in variation change of light scenarios and some occlusion problems but the main contribution of this paper is that it works for both long and short term memory.

The technique proposed in the paper [54] is the designing technique of an optimized object tracking which is used to minimize the processing time required in the object detection process while maintaining the accuracy of an occluded object in a cluttered scene. In a linear state, the Kalman Filter is applied by assuming the Gaussian distribution [55]. Different states are estimated in K.F including past, present and future time domain. Two main operations are used in the Kalman filter the prediction step which is responsible for predicting the state to obtain a priori estimation and the correction step which is responsible for the improvement of a priori estimation resulting

in the prediction step. A Kalman filter-based crop image method is proposed in which initialization is based on the first frame, then in the prediction operation, state-based prediction is done in the process model in which covariance calculation is done after this measurement error calculation is done with Kalman gain and covariance of correct operation is done. The predicted operation and correction operation is done in a loop. In object tracking implementation the Kalman predicts the targeted object's next position. Search for the object in the entire video if the object is detected, then the Kalman filter is initialized, then the relative position for the position of the object in the next frame is predicted. Now crop the next frame image at the predicted Kalman filter location, the object detection process is performed on a smaller cropped image obtained from a video sequence with the help of Kalman filter prediction.

In this paper [56] the process causes the object into different sub-regions, to discover the information of the object which is moving. This paper uses a Kalman filter for the moving objects. Two types of filters are used in Kalman cell Kalman filter and relation Kalman filter. For the object detection and tracking accurate and real-time method is used to eliminate the causes of luminescence changes. In given method a video takes as input and then convert it into the frames after this background estimation is done with the help of background estimation method with the help of frame we will find the our object after this image is subtracted from the background in binary form at the end object comes in front of us this is a very powerful technique used in the video surveillance camera.

In this paper [57] a unique work is presented with name Kalman Filter ensemble-based visual object tracker (KFebT) A different state of the art trackers are fuse together with the Kalman filter, and presents a technique which don't need training in proposed technique a confidence result is used to evaluate the existing method by using Kalman filter in fast way. Here the two main core concepts of fusion trackers should be considered the data gives as input to the fusion procedure and the response from it. For the simple integration of trackers use the tracker's result as input to the fusion method. The tracker fusion can be done with or without feedback [58]. VOT2015 dataset is used to evaluate the proposed tracker, proposed tracker shows the best result in the form of accuracy and fails less than the other methods besides this the proposed method also reaches to a small amount of failure and shows the more than 100 fps speed which the suitable value for real-time applications.

## 2.2. Concepts and Methods Related to Tracking

In this section, the main concepts that are related to tracking are discussed which we have learned from the literature review and these are very helpful for the tracking work these are basic concepts in which we have learned what is object detection? Object detector and CNN working.

### 2.2.1. Object Detection

For the purpose of tracking it is very necessary to first detect the object accurately and then track it. So, with the help of extracted layers firstly object is detected and after that, it is tracked online/offline. Object detection includes object classification, the science of object detection is about finding the object's location and its category. The object's under track is represented as a bounding box shown in Figure 2.2.

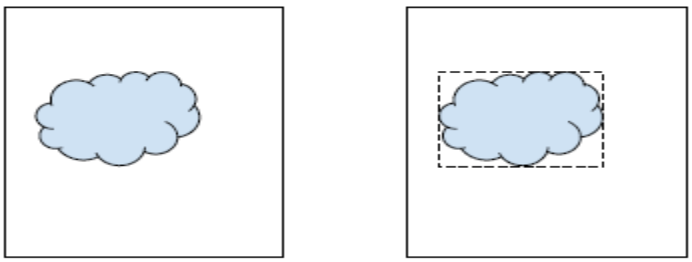


**Figure 2.2:** Result of Detection and Tracking

It has been observed that the system is relying on deep features extracted from a single layer and using a single pre-trained model is not robust enough [59]. Therefore, the alternative strategies to make the detection, robust so that tracking will be done robustly for the purpose discriminant features. Good extracted features will help in tracking the object correctly because it is detected accurately. A breakthrough has been noticed since 2012 after Alex Net [33] was popularized. Due to Single Shot Detector (SSD) the advancement is done in object detection approaches both in terms of speed and accuracy.

### 2.2.2. Object Detector

Object detectors are made up of different combination of networks the base of these networks (i.e., MobileNetv1, InceptionV2, ResNet101, Inception-ResNet, and VGG19) [60]. We extract the desired class from the detector and omit the rest of the class. Later, when it comes to updating the tracker, the framework uses the base networks to

calculate the features of the desired image. Even though in principle, the choices of the base network for object detection, and generic feature extraction can be different. By usingVGGNet-19 [44] as in [61], the features which are used are from the last convolutional fully connected layer (conv5-4). For each type of feature, a CF is to be learned.

### 2.2.3. Convolutional Neural Network

Convolutional Neural Networks models are extensively used for object detection and classification[62]. CNN has a very valuable architecture in which all neurons are connected to a set of neurons in the next layer. The basic CNN's architecture is shown in Figure 2.3. Basically, there are three fundamental concepts are used in CNN as a layer:

- ***Convolutional Layer:*** It is used to detect and extract local features in a given dataset.
- ***Pooling Layer:*** This is the significant concept in CNN, this layer is made up of max-pooling and convolution. Max-Polling is used to extract a set of maximum responses with an objective of feature reduction and provides robustness against noise and variation.
- ***Fully connected layer:*** Convolutional and pooling layers are followed by a fully connected feedforward layer, it follows the traditional fully connected layers, having input and output units.

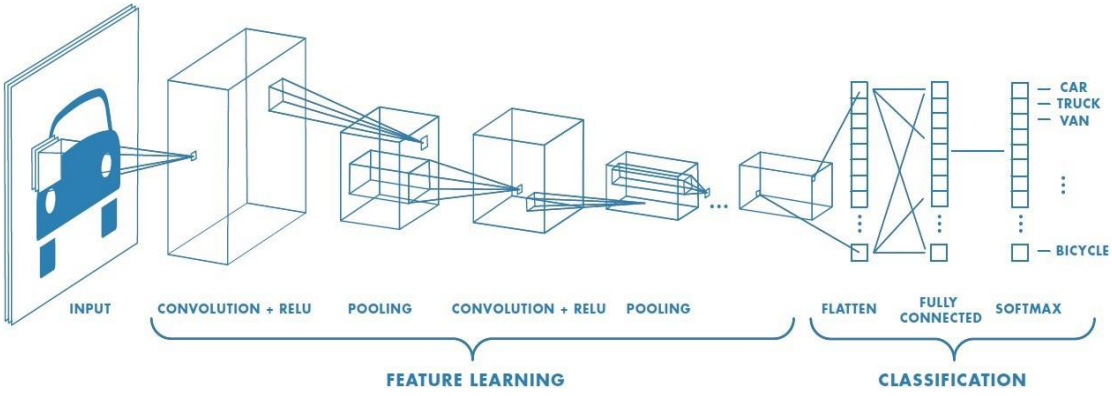



**Figure 2.3:** Convolutional Neural Network Working

### 2.2.4. Minimum Output Sum of Squared Error (MOSSE)

Minimum Output Sum of Squared Error (MOSSE) tracker, which is one of the very first approaches that apply correlation filters for the tracking problem, visual tracking initialization is done in the first frame, with the help of that frame a bounding box usually known as ground truth is generated and the tracker generates a filter representing the appearance of the target. MOSSE [38]filters can be trained through very few images. Literature shows that they are performed out of the way with respect to changes in the appearance of the foreground objects and differentiate well between background and target. The design of the filter is carried out in the Fourier domain to exploit the convolution theorem so we can define correlation as

$$G = F \otimes H^{*} \qquad (2.1)$$

Where * is the complex conjugate, $\otimes$ denotes the multiplication. G, H, and F are the Fourier transform of the Gaussian distribution G, image F and filter H. There are two components of MOSSE, the initialization, and tracking. For the initialization purpose object is selected with the help of the first few frames for a simple version of the tracker or for the complex version of the tracker. The specific object is cropped and centered to initialize the filter and after doing all this the initialized correlation filter is then correlated with a tracking window in the video to find the new location of the object.

## 2.3. Analysis of Literature Review

Literature shows that there is a lot of advancement in the VOT field by using some modern-day and conventional techniques. From literature review we come to an analysis that correlation filters are still open field of computer vision to be explored more because a huge successful results are generated with the help of correlational filters but there also some gap is still available in which the CFs are not performed well so it is decided to do work in this area to give some useful contribution to research community by playing with the existing CFs techniques and some traditional techniques so that performance can be improved, improvement is bring in term of the accuracy and performance, detailed proposed methodology is presented in section 3. But for the implementation purpose, firstly we must decide the dataset for our work from the literature review section it is observed that the dataset mostly used for the VOT

is Object Tracking Benchmark (OTB) datasets. So the data set from OTB-2013 [63] is selected by keeping in view the experimental environment (see sect 4.1). Object Tracking Benchmark (OTB) is a platform for discussing the advancements and evaluation made in the field of visual object tracking and this platform pushes the researcher to move forward in a field of visual object tracking.

## 2.4. Generating the Datasets

This section will describe the creation of datasets. We use a total of 49 different sequences; the dataset is shown in Figure 2.4 and from OTB-2013 in the evaluation of the correlation filter and the proposed technique (see sect 3.6). These all datasets contain the physical objects and were used to evaluate the techniques of object tracking.

### 2.4.1. Dataset Sequences:

The dataset [63] selected for the experimental purpose is shown in Figure 2.4 with the targeted object in the red colour bounding box and its annotation is provided by ground truth means what is the location of the desired target.

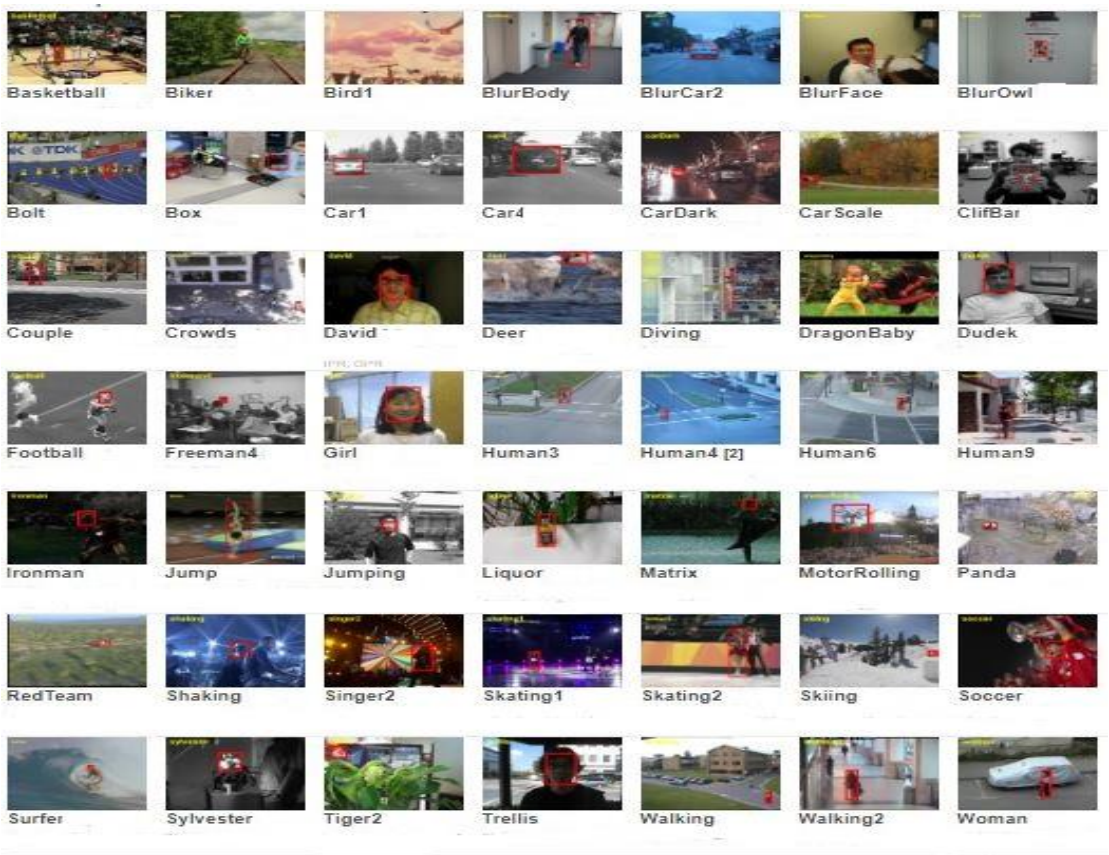



**Figure 2.4:** Selected Dataset

# Chapter 3
# Proposed Tracker

In this chapter, we present the whole method that we have used for the improvement of visual object tracking with the combination of different filters. Brief concepts of methods used are discussed in this chapter like short and long-term tracking, Kalman Filter, Kernelized Correlation filter, HOG features, and correlation and discrete Fourier transform. These concepts are related to our work and we have understood all these in detail, but they are briefly explained here in this section and how we work on this tracking task, besides this process flow of our tracker is shown in Figure 3.4.

## 3.1. Selection of Base Paper

After the analysis of literature, the latest published paper is to be selected which is related to our addressed problem of occlusion and illumination variation as a limitation. As a limitation of this paper many factors are identified to take into consideration in order to contribute to the research community from our end. So by considering all the things it is decided to select *"Adaptive Correlation Filters with Long-Term and Short-Term Memory for Object Tracking"* [53] because the method shows good results in some of the tracking issues (see Figure 1.2.) and we want to do some novel thing so that this method is able to handle most of the tracking challenges.

## 3.2. Existing Tracker

After the decision of the selection of paper now our main focus is to see the working of algorithms implemented in paper [53]. In this paper author purposed to learn multiple adaptive correlation filters with the combination of short- term memory and long-term memory of target appearance as explained in section 3.3 for the tracking purpose. They divided their learning into three main steps

- **Step 1:** Kernelized correlation filter with an aggressive learning rate for locating the target object precisely and the appropriate size of the surrounding background and feature representation is presented.
- **Step 2:** A target position is estimated with a correlation filter that uses a feature pyramid for the scale changes.

- **Step 3:** Complementary correlation filters are learned with a conservative learning rate for the longest appearance of a target. If the tracker is failed, then they apply the online detector using the SVM and detect the object in sliding window fashion.

All these approaches are used for the adoption of moving schemes average having a high learning rate for the updating of filters to be learned for handing appearance change over time. Precisely, three filters which are used and defined below

***1) Translation Filter ($A_T$):*** is used for the estimation of the target means that a target is separated from the background by enlarging the context area in window size by a ratio $r = 2.8$ which is multiplied with initial target is given by ground truth and the constant c=1.4 is a constant so the $\mathrm{A}_T$ calculated by the Eq 3.1 and for the accuracy of the target Histogram of local Intensities (HOI) is introduced as corresponding features used in the Histogram of oriented gradients (HOG) (see sect 3.4).

$$A_T = \text{target\_size} \otimes [\text{c,r}] \tag{3.1}$$

***2) Scale Filter ($A_S$)*:** is correlation is learned by regressing the feature pyramid of the target object to a one-dimensional scale space for estimating the scale variability by the scale regression model and this value is measured by given Eq 3.2.

$$A_s = \exp\left( -\frac{(s - N/2)^2}{2\sigma_o{}^2} \right) \tag{3.2}$$

***3) Long term Filter ($A_L$):*** For each tracked result, a confidence score is calculated using the long-term filter to check whether tracking failure occurs, and a confidence score is calculated by the given Eq 3.3 where a model is defined as windows size and appearance size is multiplied with Gaussian correlation.

$$A_L = \mathrm{F}(Gaussain\_correaltion \otimes Model) \tag{3.3}$$

In selected paper authors purpose is to achieve the goal that multiple correlation filters will handle the challenges of VOT[53]. (1) Appearance changes, (2) scale variation, and (3) target recovery from tracking failures by means of the dataset[63].To handle these problems they proposed a tracker having the following 6 steps.

1. Apply Translation filter **($A_T$)** for the detection of tracking objects in the selected area.
2. Scale filter **($A_S$)** to predict the scale change.
3. Long term filter **($A_L$)** is used for the detection of failure, whether it occurs or not.
4. An online detector is activated to recover the target
5. An additional module of SVM is built for the detection of failure.
6. The online detection module is responsible to recover the targeted object when a failure occurs during the target.

This algorithm performs well in the OTB2013[63] but when we run the OTB2013 [63] dataset which is also suggested in this paper and available at (http://cvlab.hanyang.ac.kr/tracker_benchmark/datasets.html)When the OTB2013 [63] is experimented then it is observed that on many of the sequences, facing different challenges of object tracking this existing algorithm is not performed well (see sect 4.3) However, the author in paper [53], claims that this algorithm is performing well favorably against the other multiple tracking algorithms like MEEM[64] and Muster[65].

### 3.3. Short-term and Long-term Tracking

A long-term tracker has the responsibility to handle the disappearance of the target and its reappearance. Few invented trackers fulfil the requirements of long-term tracking, still, there are some trackers that address this problem partially. In this paper [66] it is argued that tracker is not classified as short-term or long-term. The concept is shown in Figure 3.1 and Figure 3.2.

#### 3.3.1. Short-Term Tracking

In short, term tracking the target (green box) may move and change its appearance in the frame but it always at least partially visible. The target position is reported at each frame. The trackers do not implement target detection and do not explicitly detect occlusion. Such trackers are likely to fail on first occlusion as their representation is affected by occluding.

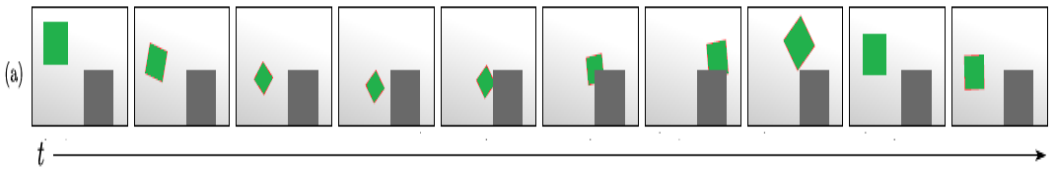



**Figure 3.1:** Short-term Tracking

### 3.3.2. Long-term Tracking

In long-term tracking, the target (green box) may disappear from the frame or it may occlude for numerous frames by the other objects (gray box) present in the frame. Due to this reason, the location of the object is not identified by the tracker. So, when the location of the target is not identified means for tracker it is not available (N/A) then the target doesn't perform any explicit re-detection function, but it uses some of its internal mechanisms for the identification and for the reporting of tracking failure.

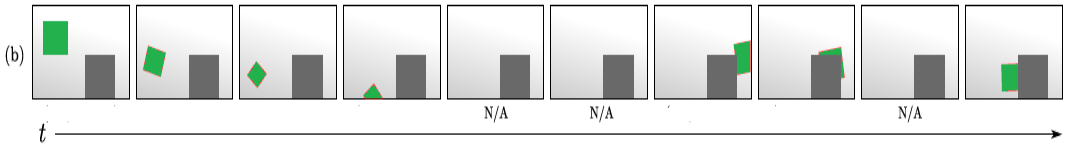



**Figure 3.2:** Long-term Tracking

### 3.4. HOG Features

HOG stands for "Histogram of Oriented Gradients" and as the name shows that it consists of image gradient orientation histograms from the image patches representing the object(s). HOG is high-level features; the feature vectors are also used to train the classifiers. On so many occasions, HOG features prove that these are too effective for the task of object description and detection especially when combine with a linear SVM classifier. They were first used by the researchers and described in the paper [42]. These gradients based descriptors perform well in high illumination and local shape deformation used to help learn discriminative correlation filters for object tracking [14, 49].

### 3.5. Correlation and Discrete Fourier Transform

Convolution, correlation, and autocorrelation are operations that can be done fast work with the application of Fast Fourier Transform (FFT). DFT of correlation of two functions is equal to the element-wise product of the Discrete Fourier Transform of one

function with the complex conjugate. We can say in this way that a Discrete Fourier Transform of a Gaussian is also a Gaussian. Smoothing with a Gaussian reduces high-frequency components of the signal.

### 3.6. Kernelized Correlation Filters

Correlation-based trackers [53, 67] achieve state of the art performance in recent benchmark evaluation[63, 68]. The idea used in these techniques regresses the circularly shifted version of the image patch to be selected to lenient the target scores (this all is done by Gaussian function and decaying from 1 to 0 when the image moves away from the target). Learning of correlation filters does not require any binary-threshold samples, it will alleviate the sampling ambiguity that adversely affects most of tracking by detection approaches, and with the help of Fast Fourier Transform (FFT) a training of Correlation filters can be done. When a new input frame is given, then a correlation response map is computed in which we crop image patch (P) with the centered positions locating in the last frame. After this response map is calculated by using the target template X in the Fourier domain and the response map is calculated by the Eq 3.3. It is the product of Gaussian correlation with motion model or appearance model (windows size and appearance, size respectively) and with the function of $\mathrm{F}(Fast\_Fourier\_Transform)$. As appearance features shows the vital role in object tracking because the separate the foreground and background in the targeted frame. So, it is important to extract features for the target objects and used them simultaneously for the best result. That is why the HOG [14, 49] and HOI features are used. HOI features bear some resemblance to the distribution field scheme[51, 52]where the statistical properties of the pixel intensities are exploited as features so these two complementary types of local statistical features are used to learn the correlation filters

### 3.7. Robust Proposed Tracker

After doing the literature review and complete understanding of the existing tracker we decide to use one of the best filters and incorporate it into the existing method so that we can achieve better results on our addressed problem. So, we decided to use the Kalman filter by using it with other filters $A_T$, $A_S$ and $A_L$ (see sect. 3.2.) to achieve the best results. The basic working principle on which Kalman filter is explained in this section. The flow chart of a whole proposed tracker is shown in Figure 3.4 which illustrates our proposed tracker named **Robust Visual Object Tracking** and we show

the concrete steps of our proposed tracker where we learn four filters from which there are correlation filters ($A_T$, $A_S$ and $A_L$). The overview of implementing the algorithm is also presented after the flow chart. The Kalman filter is recursively estimated as the state of the target of an object. KF is vastly used in different fields like economics, navigation systems and object tracking especially [69] as we are working on object tracking then we have to use KF only in the context of object tracking. It is observed that a method was presented by [70] in which properties of the Extended Kalman filter and unscented Kalman filter (UKF) are used for non-linear object tracking. For the formulation of the Kalman filter problem, we require a discrete-time linear dynamic system with additive white noise that models unpredictable disturbances. The Kalman filter is working on two main steps prediction and correction. In the motion model of moving object having some dynamic noise and some noisy observation about its position, then the KF provides an optimal estimate of its position at each time step. The optimality is only possible if all the noises are Gaussian. KF is an online process, meaning that the new observation is processed as they arrive. Filter minimizes the mean square of estimated parameters (e.g. Position, velocity). A Kalman filter applies to linear systems where the state is assumed to be distributed by a Gaussian a linear system follows the principle of superposition (law of additivity and law of homogeneity). Kalman is a region-based method for finding the region of objects in the frame. The center of the object is founded first and then uses a Kalman filter for predicting the position of it in the next frame. Following equations are used in Kalman filter for the prediction (Eq 3.4), (Eq 3.5) and estimation (Eq 3.9) of the targeted object

***Prediction:***

$$X_{k_p} = AX_{k-1} + BU_k \tag{3.4}$$

$$P_{k_p} = AP_{k-1}A^T + Q_k \tag{3.5}$$

***Measurement Input:***

$$Y_k = CX_{k_M} + Z_k \tag{3.6}$$

***Correction:***

$$K = \frac{P_{k_p}H}{HP_{k_p}H^T + R} \tag{3.7}$$

$$X_k = X_{k_p} + K[Y_k - HX_{k_p}] \quad (3.8)$$

$$P_k = (I - KH)P_{k_p} \quad (3.9)$$

The working of the Kalman filter is shown in Figure 3.3 but we brief the theoretical details with the explanation and working on the above-mentioned equations. Initial state $X_o$ and $P_o$ is given where is the state matrix in which position and velocity of the tracked object is given and $P_o$ is the sequences covariance matrix (represents the error in the estimate process) for just the first frame of a sequence. A previous state is represented by $k-1$ and is the combination of $X_{k-1}$ and $P_{k-1}$ these previous state is given to Eq 3.4 and Eq 3.5 as input to predict the new state(predicted) as $X_{k_p}$ and $P_{k_p}$ in which A and B are adaptation matrix which is used to convert input state into a new state matrix, $U_k$ is the control variable matrix used to control the different dynamics of moving target like acceleration and gravity in falling tracked object case, $Q_k$ is the process noise covariance matrix which is added in certain amount so that to $P_{k_p}$ (new process variation) not become zero or too small. Now $X_{k_p}$ and $P_{k_p}$ becomes the input to Kalman gain (Eq 3.7) and to a new state(Eq 3.8) with updated measurement Input $Y_k$ (Eq 3.6) in which $C$ is adaptation matrix, $X_{k_M}$ is the updated position and velocity of the targeted object, where $Z_k$ is the measurement noise or uncertainty. As we are observing the position of a moving object so we define a measurement matrix as $H$ and $R$ as sensor noise covariance matrix in equation in correction equations. Kalman gain (Eq 3.7) decides how much of the estimate we have to impart of measurement and how much estimate we have to impart on predicted new state Kalman gain will decide what fraction of measurement input wants to use and what fraction of predicted new state is want to use and combine them in new state (Eq 3.8).In next step, we update on process error occurred in whole process of Kalman by Eq 3.9 which later becomes the Output of updated state $X_t$ and $P_t$ shown in Figure 3.3 but current state (Eq 3.9) becomes the previous state to as input to the $X_{k-1}$ and $P_{k-1}$ this process works in circular manner till the correct prediction of object.

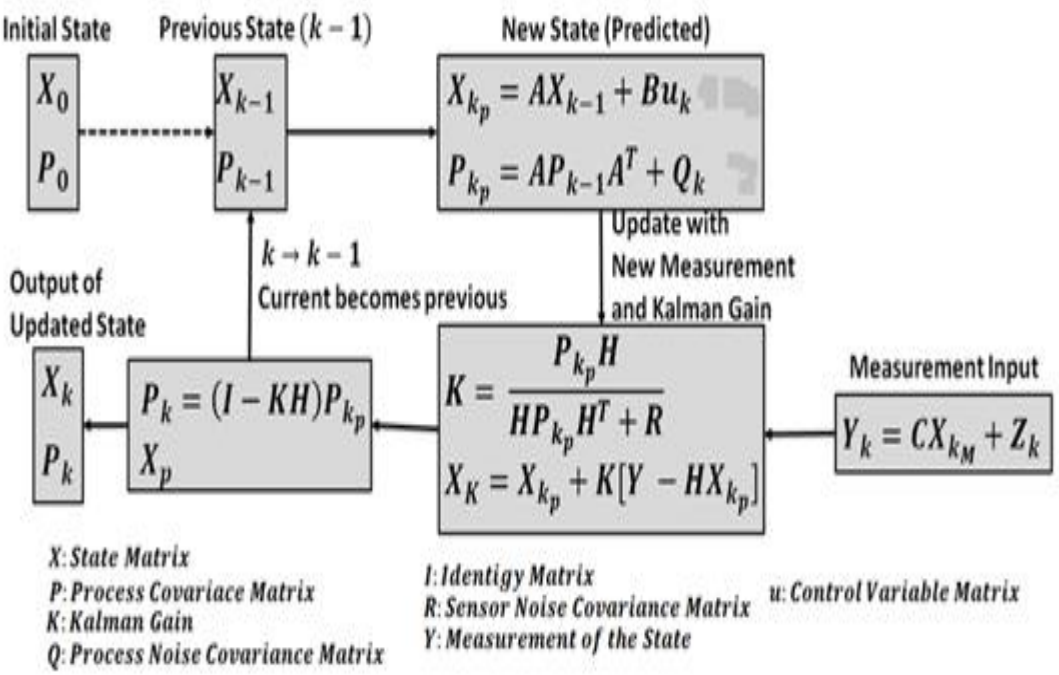



**Figure 3.3:** Process of Multi-Dimensional Kalman Filter

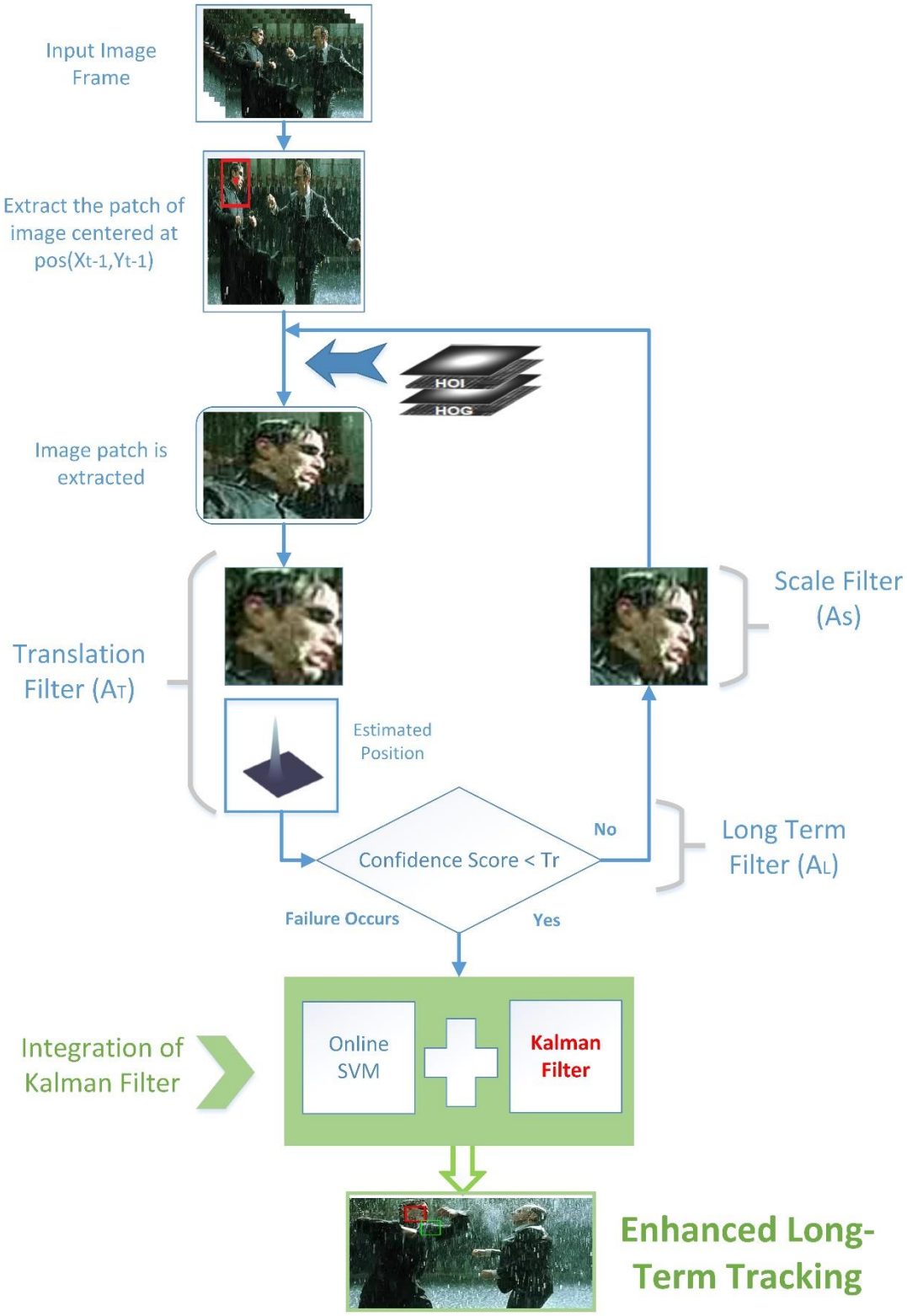



**Figure 3.4:** Proposed Tracker Working

Algorithm of the proposed tracker: Robust Visual Object Tracking

**Input:** Draw an initial bounding box $b_{t-1} = (X_{t-1}, Y_{t-1}, W_{t-1}, H_{t-1})$

**Output:** Estimated Bounding box $= (X_t, Y_t, W_t, H_t)$

1. Repeat

**2.** From initial frame crop the patch **P** with centered position $(X_{t-1}, Y_{t-1})$ and Do feature extraction with HOG and HOI features

**3.** Compute Translation filter $A_T$ (P) and locate target potion $(X_t, Y_t)$;

**4.** Scale pyramid is constructed for **P'** around the target position $(X_t, Y_t)$ and Compute Scale filter $A_S$ (P')

//Re-detection

**5.** If $Max(A_L(P)) < T_r$ then

Activate the detection module h

For each state of P do Kalman Filter (K.F)

**6.** End

// Model Update

**7.** Update $A_T$, $A_S$ and KF

**8** Until end of image sequences;

Now we must implement this proposed algorithm to validate the performance of it that how much it is competent to resolve the addressed problem. We consider all the results in quantitative and qualitative form, first, we must set the environment to perform the experiments, and then we will decide about the performance metric and perform the experiments to evaluate the existing and proposed tracker.

# Chapter 4
# Experiments and Results

In this chapter, the quantitative and qualitative result of existing and proposed tracker is presented. The chapter also discusses the experimental environment and performance metrics. The problem in existing tracker with solution is given with, the comparison of both to evaluate our proposed tracker.

## 4.1. Experimental Environment

The existing and the proposed algorithm is implemented in MATLAB 2019a by using some libraries of Open CV v.4.1.0 and Python v.3.7.4 on a PC having Intel (R) Core™ i3 @ 2.10 GHz CPU and 8 Gigabytes of RAM. All the other competing tracker codes are obtained from the respective authors' websites and implemented on the same machine for consistency. We use Visual Studio 2015 for some Mex files generated from the C++ code to be implemented in MATLAB.

## 4.2. Performance Metric

Mostly trackers are evaluated on precision of tracker with time it consumes so we interested in two performance parameters to check the tracking failure similarly we again check the performance of our proposed algorithm on the following parameter.

**(1) Precision:** Result of estimated bounding box around the object in all frames without failure. For the calculation of precision, we need a distance of position value and the ground truth value which is calculated by the Eq 4.1 and graph will be plotted at the output points of $Dis$ which will work in a loop until the number of elements present in a cell.

$$Dis = \sqrt{(position_1 - ground\_truth_2)^2 + (position_2 - ground\_truth_2)^2} \quad 4.1$$

**(2) Frames Per Second (FPS):** It is the execution speed to make the tracker work in real-time it is calculated by the given equation 4.2.

$$FPS = \frac{\text{Total\_Images}}{\text{Time}} \quad 4.2$$

### 4.3. Quantitative analysis of Existing Tracker

When we will do the quantitative analysis of the existing tracker and we came to know through the metrics of precision and frame per second (FPS) the existing algorithm is unable to show good results (see table 4.1) on a few tracking challenges. We show the performance based on precision value that how much tracker is robust to track the object. In the given dataset some video sequences show the perfect precision which is equal to 1.00 means 100%, few are below the 1.00, including the blur motion or due to fast motion, but many of them are below 0.50. We only consider those datasets whose precision is below the 50%, which we considered as a gap or a failure of this tracker and 50% is threshold all datasets below this threshold is represented in bold text and overall performance of OTB2013[63] on the existing technique with precision rate of 20 pixels and tracking speed in frame per second (FPS) is shown in table 4.1. Now tracker needs some improvements so that it can perform well on the datasets which have precision below than 50% so that it can track an object accurately and efficiently for the improvement purpose we only select those datasets which are related to our addressed problem i.e., occlusion and illumination variation (High and low both). Selected datasets are represented with bold and underlined text which we consider for the improvement with the help of recently proposed and traditional approaches to give some contribution to the research community.

**Table 4.1:** Quantitative Results and Overall Performance of OTB2013

| Dataset Name | Precision at 20px | Time: Frame Per Second (FPS) | Description |
|---|---|---|---|
| Basketball | 1 | 19.05 | Tracker performs well on this data set. |
| **Biker** | **0.458** | **22.02** | Tracker loses its ToI when biker changes the position of his face and move fastly. This is a challenge of fast object motion. |
| **Bird1** | **0.346** | **16.15** | Tracker loses its ToI when the targeted birds go in clouds and faces the occlusion due to clouds. |
| Blur body | 0.967 | 6.49 | The tracker shows better results on the Blur body data set. |
| Blurcar2 | 0.998 | 5.58 | The tracker shows better results in a Blurcar2 data set. |

| | | | |
|---|---|---|---|
| Blur face | 1.000 | 5.95 | Tracker performs well on the Blur face data set. |
| Blur owl | 0.891 | 4.02 | Tracker loses its position when the camera moves quickly but still, it shows some better results. |
| Bolt | 1.000 | 25.20 | Tracker performs well on the Bolt data set. |
| Box | 0.680 | 3.69 | Tracker loses its position when the box moves left and right that is why it does not show good results, but the results are near to 70%. |
| **Car1** | **0.438** | **8.04** | Tracker loses its ToI when shadows of trees appear and there is low illumination in the path followed by the object but it's precision it less than 0.50. |
| Car4 | 0.989 | 6.01 | The tracker shows better results on car4 data because precision is approximately near to 99%. |
| Cardark | 1.000 | 16.10 | Tracker performs well on the Cardark data set. |
| Clifbar | 0.939 | 12.16 | Tracker loses its position on the target when the object moves quickly or come near to the camera, but still, it shows some better result. |
| Couple | 0.571 | 18.40 | Tracker loses its target when the camera starts shaking frequently. |
| Crowds | 1.000 | 25.24 | Tracker performs well on Crowds data set. |
| David | 1.000 | 6.63 | Tracker performs well on David's data set and shows accurate results. |
| Deer | 0.817 | 4.25 | Trackers change their position and lose the target i.e., deer when deer start moving up and down frequently. |
| Diving | 0.753 | 20.02 | Tracker changes its position to lose the ToI when the diving object changes its body size. |

| | | | |
|---|---|---|---|
| Dragon baby | 0.549 | 9.15 | Tracker loses its position when objects move quickly. |
| Dudek | 0.907 | 4.50 | Dudek performs well, but when the target changes its appearance by removing the glasses, then Tracker gives less precision as compared to the wear glasses. |
| Football | 1.000 | 9.95 | Tracker performs well on the Football data set. |
| Freeman4 | 0.951 | 31.02 | Tracker loses its position when the newspaper comes in front of man for short intervals but still shows precision approximately 95%. |
| Girl | 1.000 | 11.78 | Tracker performs well on the Girl data set. |
| **Human3** | **0.006** | **10.1** | Tracker loses its ToI when there are similar peoples in a video and the camera is shaking it is unable to track humans. |
| Human4 | 0.852 | 21.03 | Tracker loses its position when a person moves away from the camera at the end of the video sequence. |
| **Human6** | **0.285** | **27. 00** | Tracker loses its position when the camera moves towards (zoom) the target. |
| Human9 | 0. 774 | 15.82 | Tracker loses its position when the camera starts shaking frequently |
| **Ironman** | **0.145** | **11.81** | Tracker loses its ToI due to the reason for the fast motion of the object. |
| **Jump** | **0.041** | **6.81** | Tracker loses its position when the object quickly changes its position in gymnastic and the result is below 50%. |
| Jumping | 0.978 | 9.95 | Tracker loses its position at the start, but after a small interval, it again gains its position which shows that the tracker is adaptive. |

| | | | |
|---|---|---|---|
| Liquor | 0.789 | 12.39 | Tracker loses its position when another similar object (bottles similar to the target bottle) brings near to it. |
| **Matrix** | **0.360** | **16.35** | Tracker loses its position when a targeted object faces a high illumination due to light. |
| **MotorRolling** | **0.043** | **6.85** | Tracker moves out of the frame when the motorbike changes its position in rolling it is due to the fast motion of motorbike. |
| Panda | 0.500 | 20.77 | Tracker loses its position when an object moves away from the camera, and it stops tracking the object when the object comes towards the camera. |
| RedTeam | 1.000 | 25.12 | Tracker performs well on RedTeam data set. |
| Shaking | 0.984 | 6.35 | When an object moves fast during singing, it loses for its ToI for some time, but after that, it tracks better. |
| Singer2 | 0.973 | 5.05 | Tracker lose their potion when the object slightly turns away from during singing. |
| Skating1 | 1.000 | 15.93 | The tracker shows good results in a Skating1 data set. |
| Skating2 | 0.983 | 13.68 | Tracker loses its position due to fast motion but still shows a better result. |
| **Skiing** | **0.136** | **10.38** | Tracker hasn't performed in fast motion when an object jumps with high speed. |
| **Soccer** | **0.151** | **5.08** | Tracker loses its position when the object moves quickly up and down and the sequence gets blurred. |
| Surfer | 0.979 | 24.52 | Tracker performs well on this dataset in which surfing is done and shows good results. |

| | | | |
|---|---|---|---|
| Sylvester | 0.975 | 7.80 | In Sylvester the bear toy with hand of human is being tracked, it shows good results. |
| Tiger2 | 0.693 | 6.23 | Tracker loses its ToI i-e tiger2 toy when it moves quickly. |
| Trellis | 1.000 | 3.50 | Tracker performs well on Trellis (person) data set. |
| Walking | 1.000 | 22.13 | Tracker performs well on walking data set it's accurately showing 100% result. |
| **Walking2** | **0.404** | **14.51** | Tracker loses its ToI when another person occulted the ToI by coming in front of the tracking object. |
| Women | 0.940 | 20.56 | Tracker loses its position when the camera moves towards (zoom) the selected target. |

## 4.4. Precision plot of existing tracker

We consider only 12 datasets from table 4.1 in which the precision threshold is less than 0.50 means less than 50%. As we are interested in occlusion and the illumination variation problem so we selected those sequences which are not working on these problems shown from description column in Table 4.1 shows that only four datasets are those which is related to our addressed problem so we show the graphs for quantitative analysis and more understanding of the problem that on which dataset existing algorithm does not perform well.

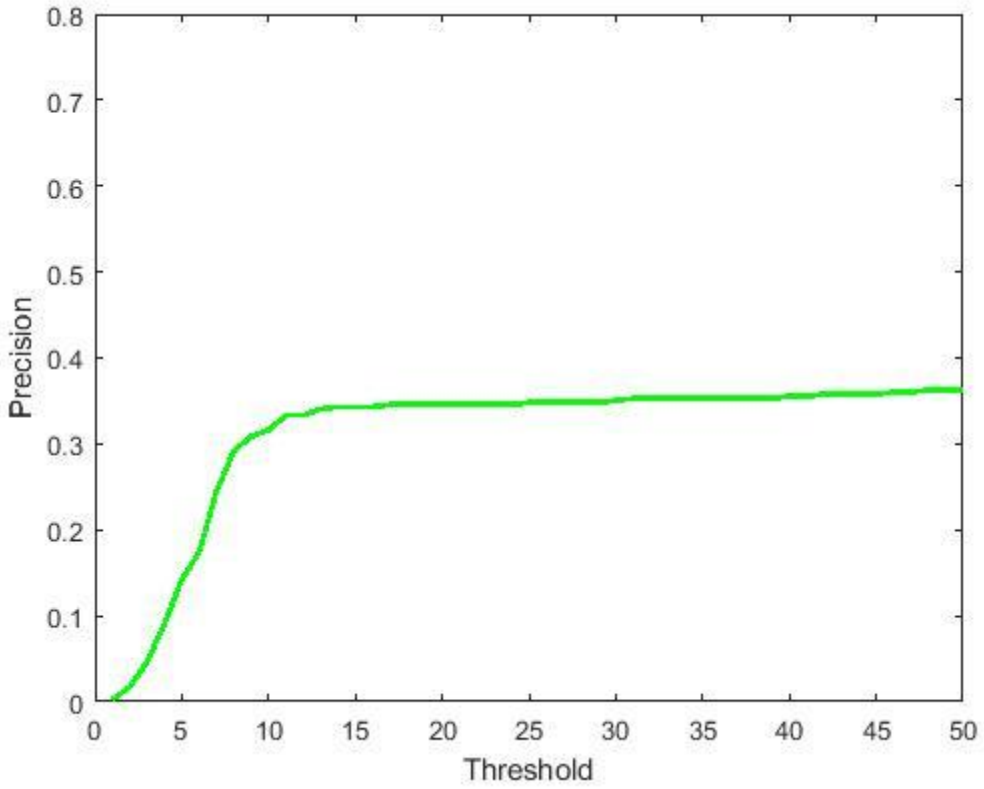



**Figure 4.1:** Existing Tracker Precision Plot of Bird1

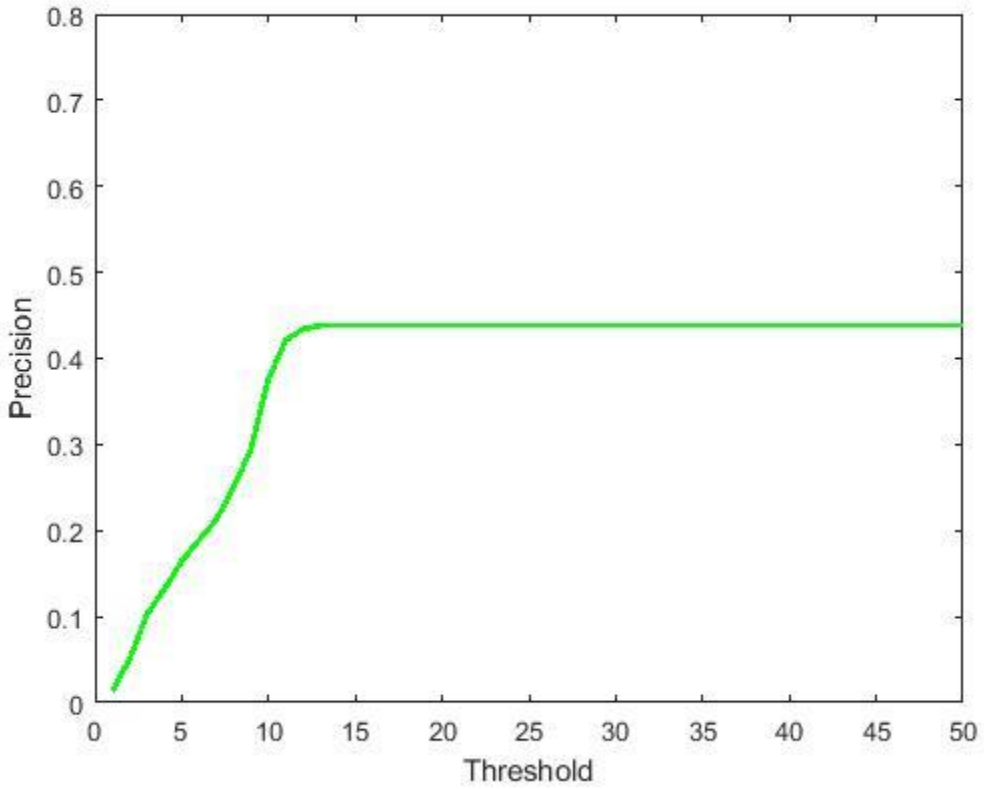



**Figure 4.2:** Existing Tracker Precision Plot of Car1

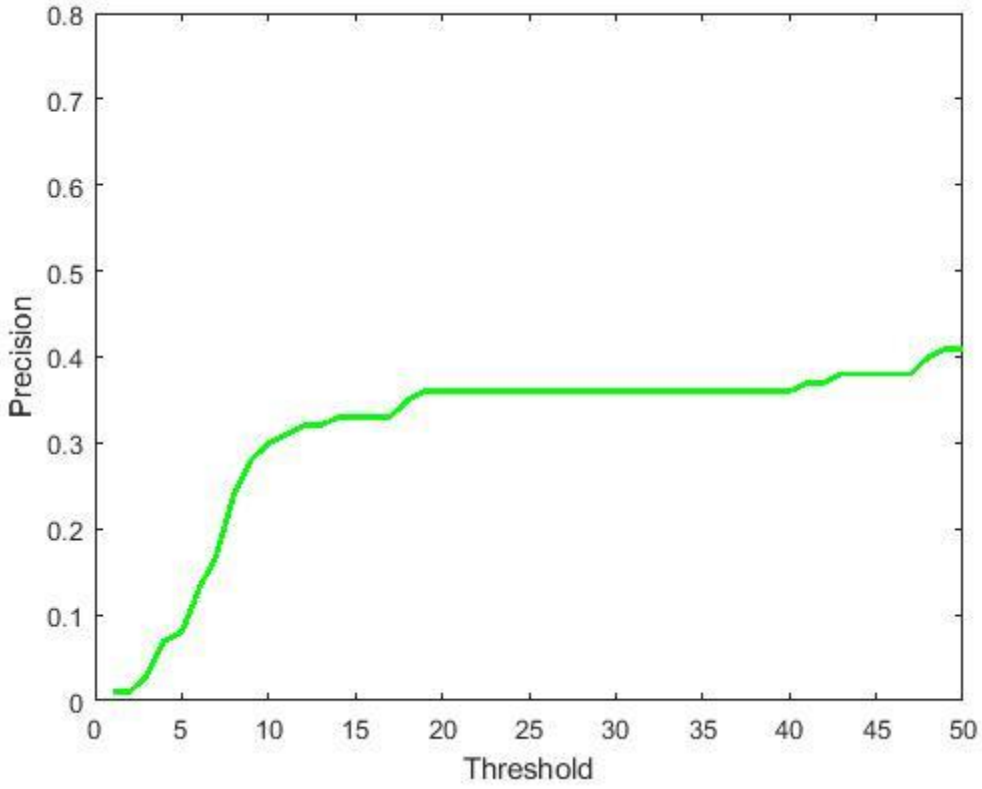



**Figure 4.3:** Existing Tracker Precision Plot of Matrix

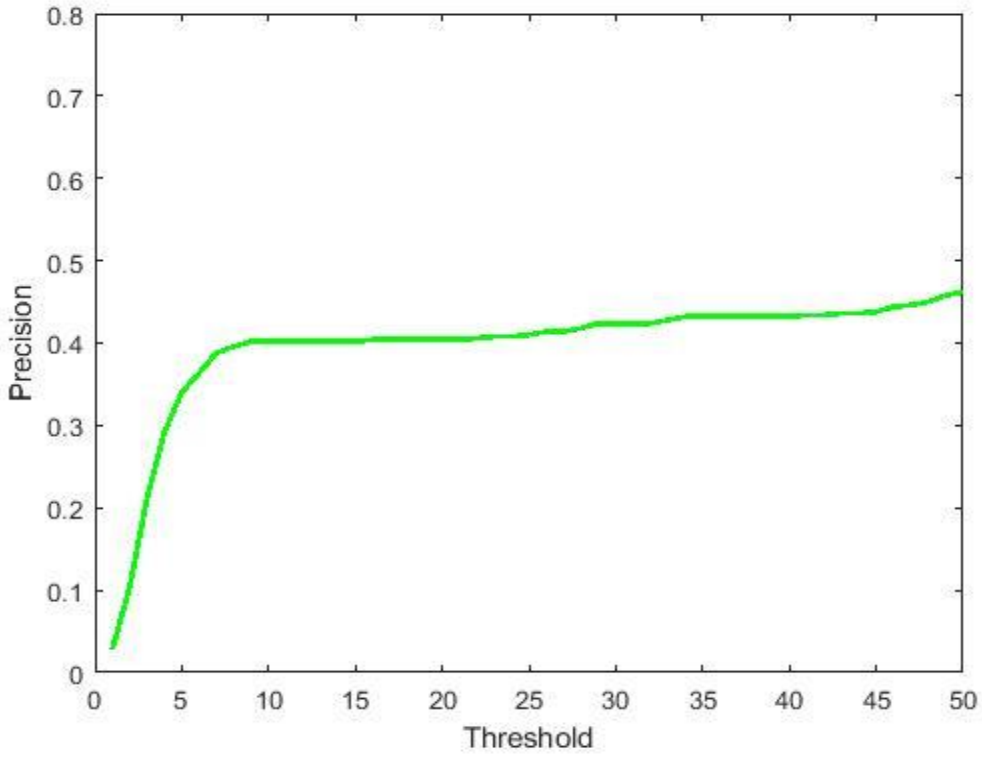



**Figure 4.4:** Existing Tracker Precision Plot of Walking2

## 4.5. Solution of Problem

For the above-mentioned problem, we give a solution in which we proposed a new tracker (see Figure 3.4) in which we use the Kalman filter with the other three correlation filters to obtain a good result by addressing our problems. In proposed tracker input is given in the form of image frame from which patch **P** is extracted with centered position after that translation filter $\mathbf{A_T}$ is applied separates the target object from the background and we came to know the estimated target position and after this a scale filter $\mathbf{A_S}$ is applied to predict the scale changes. When these filters are applied, we came to know about the failure with the help of $\mathbf{A_L}$ to check that failures occur or not (where the confidence score is less than a certain threshold). It is the scenario where the tracker loses its target due to too many reasons for tracking challenges as shown in Figure 1.2. In return for this failure, an online detection module is activated using Support Vector Machine (SVM) and **Kalman filter** which conservatively learns the target to capture the target appearance over a long span. Updating the $\mathbf{A_T}$, $\mathbf{A_S}$ and $\mathbf{K_F}$ in each frame until the image of sequences is not ended.

## 4.6. Qualitative and Quantitative Comparison of Both Trackers

By using the solution provided (sect 4.5) we produce the quantitative and qualitative results of both trackers for better understating of our proposed tracker that what we are investigating, what type of the problem we addressed and make the observation of the data related to the concept that we have proposed earlier for this purpose. Figure 4.5 shows the previous frame number f and the next frame number in which tracker loses its target on which the selected datasets are failing, that is the point where failure occurs a Bird1 dataset is failed on F #183 when the occlusion due to clouds occurs, the walking2 failed at F # 214 due to occlusion of another person in a video sequence, car1 at F # 140 due to low illumination and Matrix fails at F #36 due to high illumination of light. Now we apply the proposed tracker to the data set shown in Figure 4.5.


(a) Bird1 (b) walking2


(c) Car1 (d) Matrix

**Figure 4.5:** Existing Tracker failure

Now we will apply our proposed technique to datasets and capture the few frames after some intervals to check the robustness of our tracker to check the proposed tracker is still tracking the object when the existing tracker loses its tracking in the frame and Kalman filter is applied, after that we plot the precision graphs to validate the working of our tracker with the help of video frames and graphs we show that our tracker is achieved favorably good results after the integration of Kalman filter to the existing technique. Here it's important to mention one thing that existing tracker is represented by the green bounding box and all its precision graphs are also in green color while our tracker is represented by a red bounding box (after the Kalman update see Eq 3.9), and the black bounding box is the estimated position in next frame see equation 3.4 and 3.5. All graphs plotted for this should be in red color to see the clear difference. The video frame of results is shown in Figure 4.6.

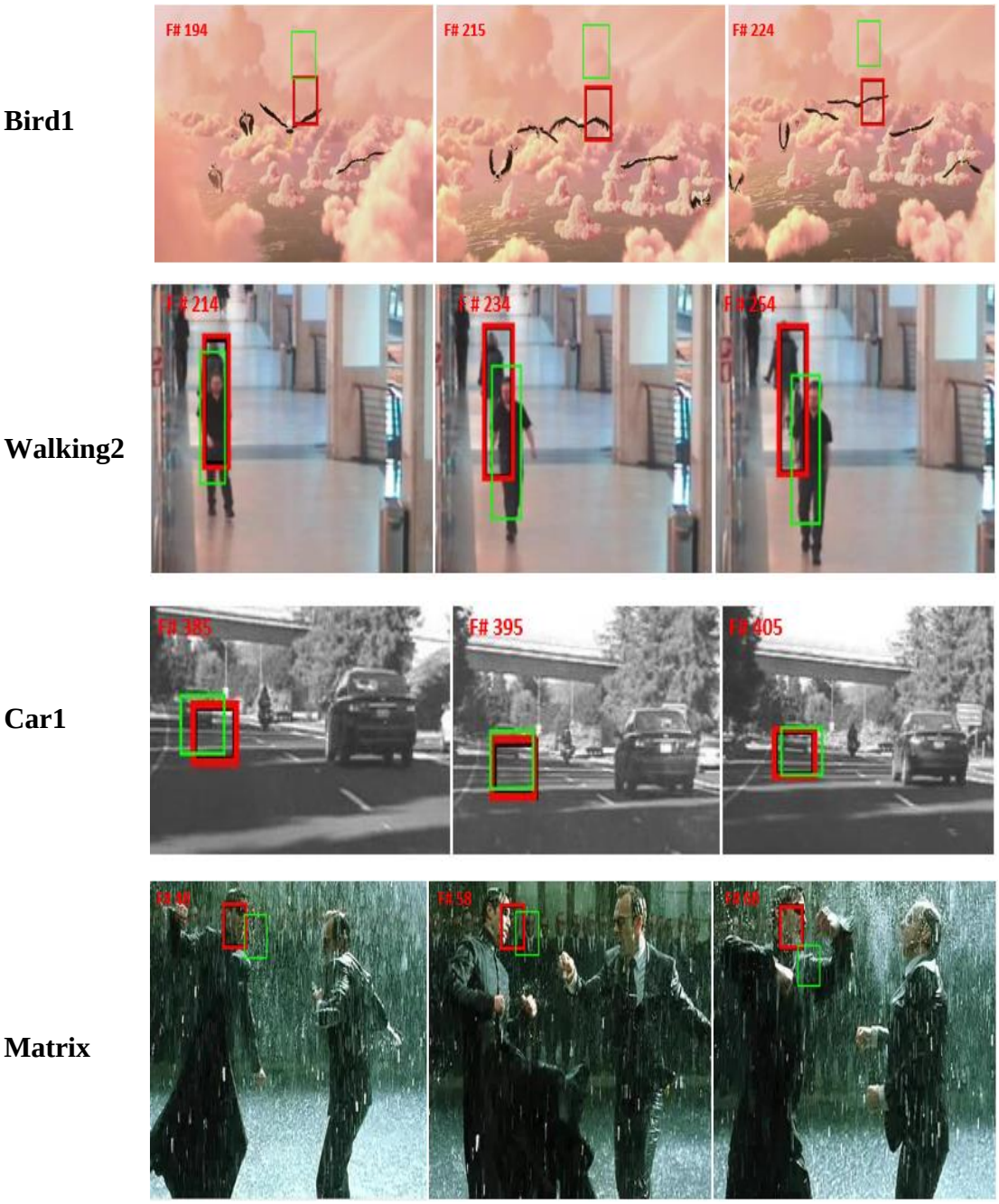



**Figure 4.6:** Qualitative Comparison of Proposed Tracker and Existing Tracker

Tracking results on four challenge sequences of our addressed problem which is shown in Figure 4.5 from first to the last row. Now precision graphs are generated to validate the performance of the proposed tracker. The comparison can also be done by considering Figure 4.6. Table 4.2 also shows the quantitative comparison with the improvement percentage of existing and proposed trackers.

**Table 4.2:** Comparison table of Selected dataset on existing and Proposed Tracker

| Problem Addressed | Dataset Name | Existing Tracker[53] | | Proposed Tracker | | Improvement |
|---|---|---|---|---|---|---|
| | | Precision | FPS | Precision | FPS | % |
| Occlusion | **Bird1** | 0.346 | 16.15 | **0.593** | 16.15 | 24.7% |
| | **Walking2** | 0.404 | 14.51 | **0.798** | 12.38 | 39.4% |
| Illumination | **Car1** | 0.438 | 8.04 | **0.594** | 7.48 | 15.6% |
| | **Matrix** | 0.360 | 16.35 | **0.640** | 11.41 | 28% |

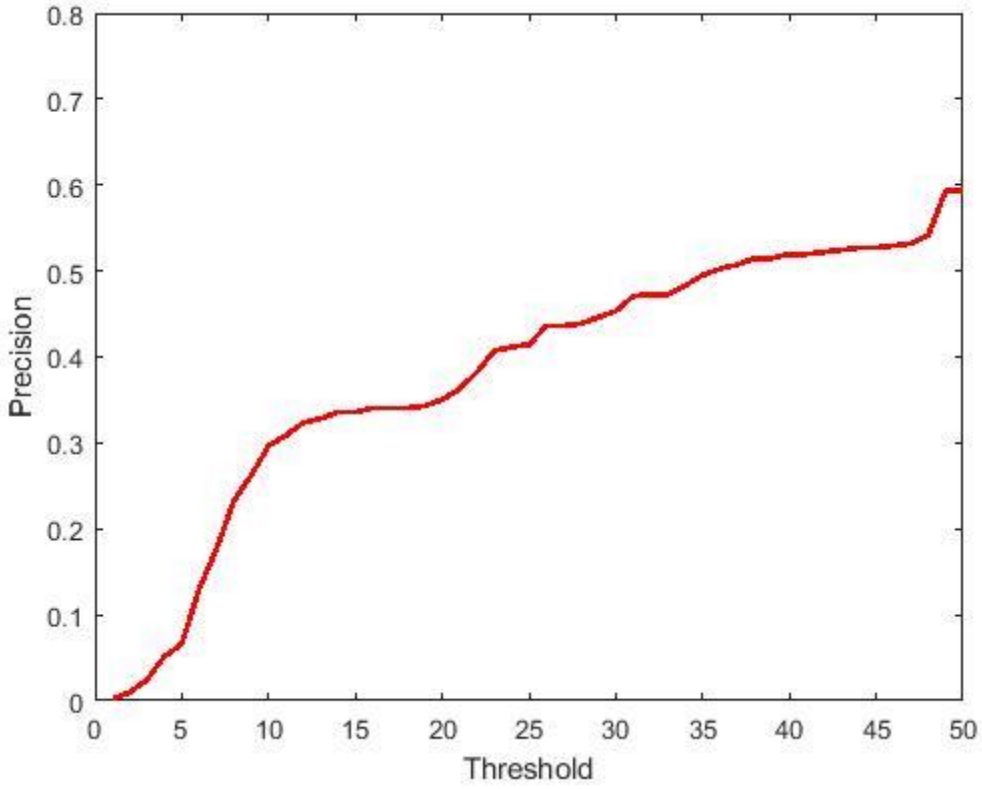



**Figure 4.7:** Proposed Tracker Precision plot of Bird1

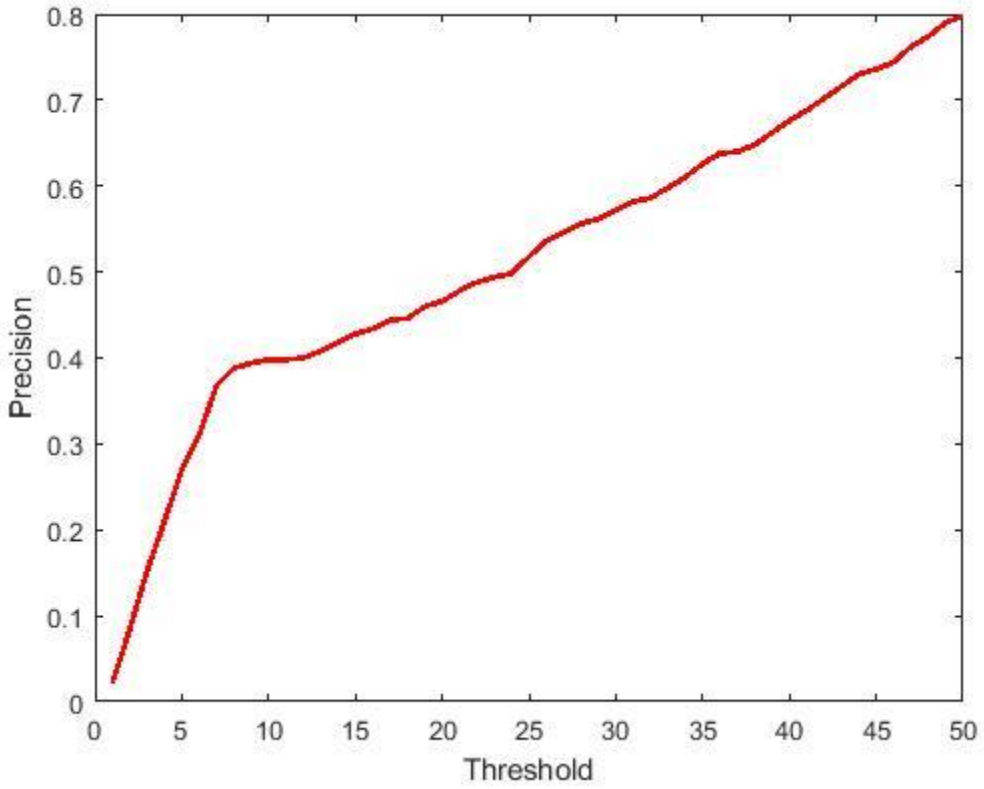



**Figure 4.8:** Proposed Tracker Precision plot of walking2

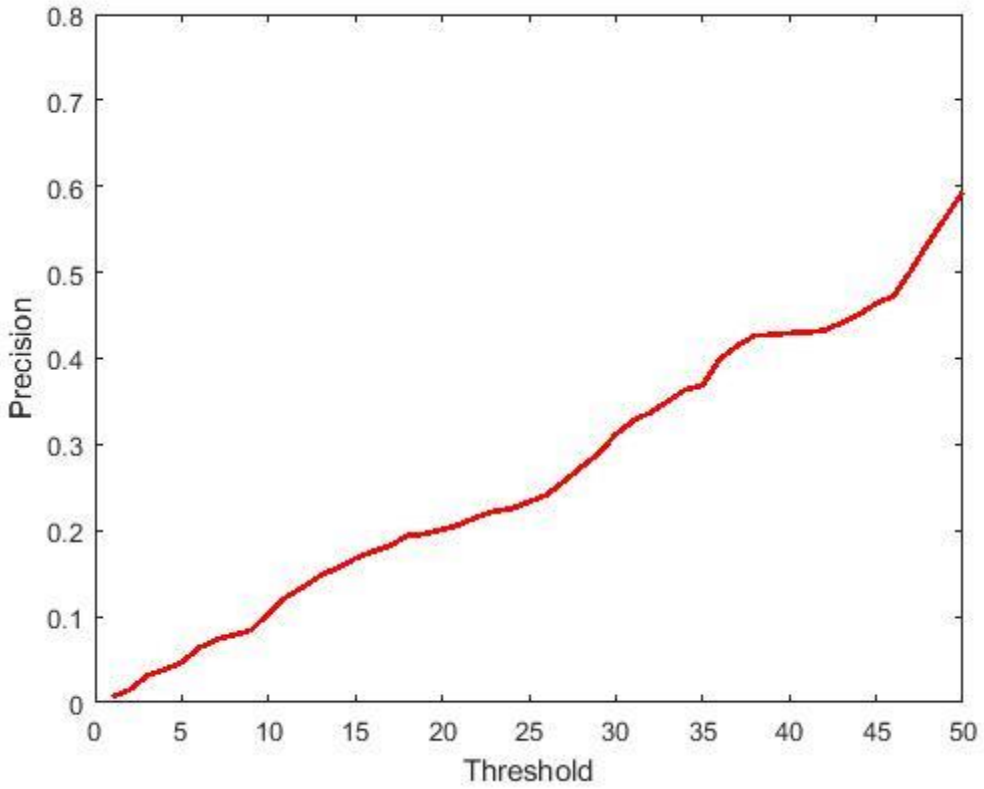



**Figure 4.9:** Proposed Tracker Precision plot of Carl

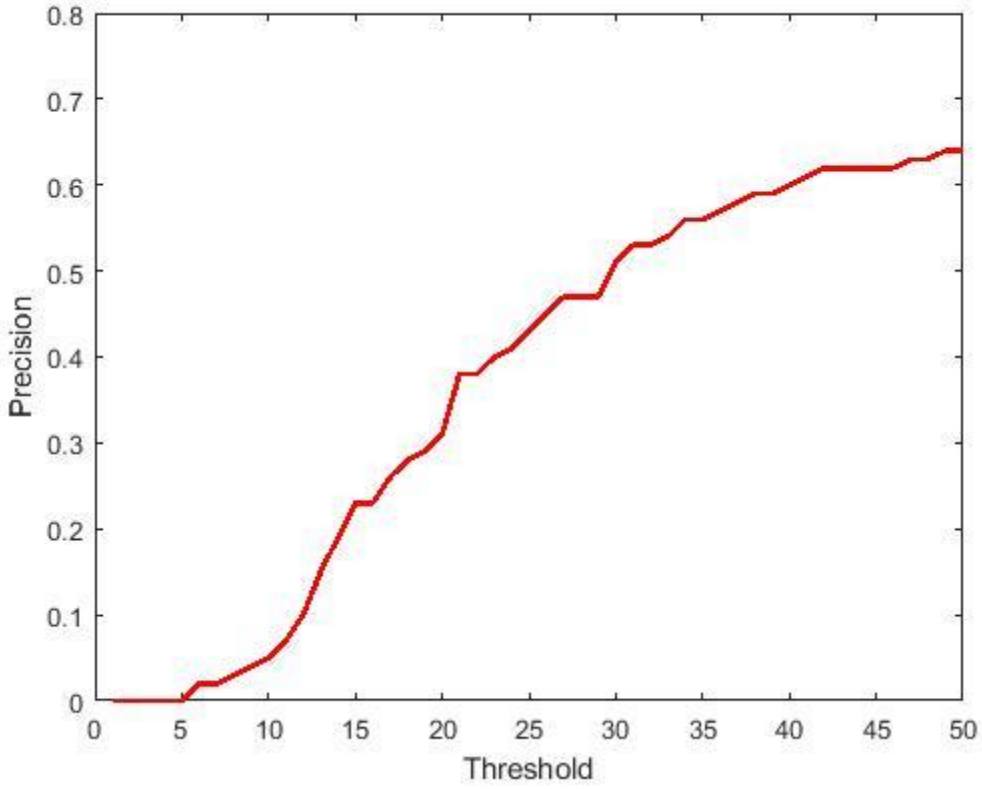



**Figure 4.10:** Proposed Tracker Precision plot of Matrix

We have improved the selected datasets highlighted in (sect 4.4) whose precision is less than 0.50 and related to addressing problems and we bring improvement in the selected dataset with some good results. We can see that improvement is done by using the Kalman filter with other filters ($A_T$, $A_S$ and $A_L$). The improvements can be easily seen in Table 4.2 that we bring the improvement in Bird1 of 24.7% by improving the precision from 0.36 to 0.593, for Walking2 39.4% improvement is done by improving the precision from 0.404 to 0.798 in the occlusion problem Similarly for the illumination variation problem we have two dataset Car1 in which we bring the improvement of 15.6% and in terms of precision moves from 0.438 to 0.594, In Matrix the 28% improvement is done in which the precision moves from 0.360 to 0.640 the improvements by our tracker is shown in bold text with red color.

# Chapter 5
# Conclusion and Future Work

## 5.1. Conclusion

In this work we address the problem of single object visual tracking by considering the two tracking challenges i.e., occlusion and illumination (high and low both) and brings improvement in the result. The goal is to achieve a favorable improved result in comparison to using some recently proposed and new proposed method. We are using adaptive correlation filters, including the Long-term filter ($\mathbf{A_L}$**),** Translation Filter **(**$\mathbf{A_T}$**)** and Scale Filter **(**$\mathbf{A_S}$**)**. In which image patch of the selected target is extracted with the cantered positions and the features are extracted with the help of HOG and HOI and the object is being tracked. But unfortunately this algorithm is not performing well on some of the tracking challenges (see table 4.1) and not shows desired results on selected sequence, we will decide to improve the results of those datasets on which the existing algorithm is not working so we do a novel thing in this so we integrate the Kalman filter to the existing algorithm after complete understanding of an existing algorithm although it's not a simple task to do but we do it successfully. We apply the Kalman to that part of the algorithm where the existing trackers fail to track the target object. When we integrate Kalman to it and performed the experiments on same datasets which were used by existing algorithms, with the help of quantitative (see table 4.2) and qualitative results (see Figure 4.6) we came to the conclusion that our proposed tracker provides the improved results above the threshold which we decide earlier to improve those data sets whose precision values are less than 50% and we can say that if proposed tracker (see sect 3.7) are used on tracking challenges of occlusion and illumination (high and low both) a good results are obtained.

## 5.2. Future work

We have improved the those datasets which relate to our problem of interest and whose precision is lower than the 0.50 and we only address the two problems (occlusion and illumination variation) of VOT but in the future those datasets are also begin addressed which follows the fast-moving object motion challenge and their precision is less than 0.50. Precision of giving dataset with brief descriptions and VOT challenges is quantitatively given in table 4.1 for the easiness of future work. A devise strategy like

increasing the window size around the target object will also improve the problem of fast-moving objects. We can formulate some other robust conventional technique with correlation filters for the improvement. Currently, a combination of the two algorithms is used in experiments. Other revolutionary algorithms such as Minimum Output Sum of Squared Error (MOSSE), Multiple Experts using Entropy Minimization (MEEM) or Discriminative scale space tracking (DSST) are also being used for the improvements of remaining VOT challenges. It is notable that fast-moving objects are also important tracking challenges to be considered for surveillance systems for such requirements as fruitful work is done by using the above mention techniques. In the future, work is done for improvements of VOT challenges and sub-field of computer vision grows with tremendous speed to facilitate every individual who is associated with its application. But for now, we leave the aforementioned discussion as some thoughts for the future development of this thesis.

# References


1. Ali, A., et al., *Visual object tracking—classical and contemporary approaches.* Frontiers of Computer Science, 2016. **10**(1): p. 167-188.
2. Kim, I.S., et al., *Intelligent visual surveillance—A survey.* International Journal of Control, Automation and Systems, 2010. **8**(5): p. 926-939.
3. GDANSK, K., *Deliverable 2.1–Review of existing smart video surveillance systems capable of being integrated with ADDPRIV. ADDPRIV consortium*. 2011.
4. He, Q., et al., *TA-2, a thrombin-like enzyme from the Chinese white-lipped green pitviper (Trimeresurus albolabris): isolation, biochemical and biological characterization.* Blood Coagulation & Fibrinolysis, 2012. **23**(5): p. 445-453.
5. Lee, J., et al. *Strategies of path-planning for a UAV to track a ground vehicle*. in *AINS Conference*. 2003.
6. Ahmed, J., et al. *A vision-based system for a UGV to handle a road intersection.* in *Proceedings of the National Conference on Artificial Intelligence*. 2007. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.
7. Sakagami, Y., et al. *The intelligent ASIMO: System overview and integration*. in *IEEE/RSJ international conference on intelligent robots and systems*. 2002. IEEE.
8. Mondragón, I.F., et al. *Visual model feature tracking for UAV control*. in *2007 IEEE International Symposium on Intelligent Signal Processing*. 2007. IEEE.
9. Amini, A.A., et al. *Non-rigid motion models for tracking the left-ventricular wall*. in *Biennial International Conference on Information Processing in Medical Imaging*. 1991. Springer.
10. Vasconcelos, M.J.M., et al., *Using statistical deformable models to reconstruct vocal tract shape from magnetic resonance images.* Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 2010. **224**(10): p. 1153-1163.
11. Mistry, P. and P. Maes. *SixthSense: a wearable gestural interface*. in *ACM SIGGRAPH ASIA 2009 Sketches*. 2009. ACM.
12. Zhu, Z. and Q. Ji. *Eye gaze tracking under natural head movements*. in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. 2005. IEEE.
13. Fiaz, M., et al., *Handcrafted and Deep Trackers: Recent Visual Object Tracking Approaches and Trends.* ACM Computing Surveys (CSUR), 2019. **52**(2): p. 43.
14. Henriques, J.F., et al., *High-speed tracking with kernelized correlation filters.* IEEE transactions on pattern analysis and machine intelligence, 2014. **37**(3): p. 583-596.
15. Schroff, F., D. Kalenichenko, and J. Philbin. *Facenet: A unified embedding for face recognition and clustering*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
16. Nam, H. and B. Han. *Learning multi-domain convolutional neural networks for visual tracking*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
17. Zhang, K., et al., *Robust visual tracking via convolutional networks without training.* IEEE Transactions on Image Processing, 2016. **25**(4): p. 1779-1792.

18. Gao, C., et al., *Robust visual tracking using exemplar-based detectors.* IEEE Transactions on Circuits and Systems for Video Technology, 2015. **27**(2): p. 300-312.
19. Xu, C., et al., *Robust visual tracking via online multiple instance learning with Fisher information.* Pattern Recognition, 2015. **48**(12): p. 3917-3926.
20. Yang, H., S. Qu, and Z. Zheng, *Visual tracking via online discriminative multiple instance metric learning.* Multimedia Tools and Applications, 2018. **77**(4): p. 4113-4131.
21. Held, D., S. Thrun, and S. Savarese. *Learning to track at 100 fps with deep regression networks*. in *European Conference on Computer Vision*. 2016. Springer.
22. Wang, J., et al., *Two-level superpixel and feedback based visual object tracking.* Neurocomputing, 2017. **267**: p. 581-596.
23. Huang, W., et al. *Structural superpixel descriptor for visual tracking*. in *2017 International Joint Conference on Neural Networks (IJCNN)*. 2017. IEEE.
24. Filali, I., M.S. Allili, and N. Benblidia, *Multi-scale salient object detection using graph ranking and global–local saliency refinement.* Signal Processing: Image Communication, 2016. **47**: p. 380-401.
25. Du, D., et al., *Geometric hypergraph learning for visual tracking.* IEEE transactions on cybernetics, 2016. **47**(12): p. 4182-4195.
26. Yao, R., et al., *Part-based robust tracking using online latent structured learning.* IEEE Transactions on Circuits and Systems for Video Technology, 2016. **27**(6): p. 1235-1248.
27. Wang, J., et al. *Part-based multi-graph ranking for visual tracking*. in *2016 IEEE International Conference on Image Processing (ICIP)*. 2016. IEEE.
28. Zhang, T., et al. *Structural sparse tracking*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
29. Zhang, T., et al., *Robust visual tracking via exclusive context modeling.* IEEE transactions on cybernetics, 2015. **46**(1): p. 51-63.
30. Yi, Y., Y. Cheng, and C. Xu, *Visual tracking based on hierarchical framework and sparse representation.* Multimedia Tools and Applications, 2018. **77**(13): p. 16267-16289.
31. Lee, J., et al., *Globally optimal object tracking with fully convolutional networks.* arXiv preprint arXiv:1612.08274, 2016.
32. Li, B., et al. *Siamrpn++: Evolution of siamese visual tracking with very deep networks*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.
33. Krizhevsky, A., I. Sutskever, and G.E. Hinton. *Imagenet classification with deep convolutional neural networks*. in *Advances in neural information processing systems*. 2012.
34. Feng, W., et al., *Dynamic Saliency-Aware Regularization for Correlation Filter-Based Object Tracking.* IEEE Transactions on Image Processing, 2019. **28**(7): p. 3232-3245.
35. Qin, Y., et al. *Saliency detection via cellular automata*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.
36. Kart, U., et al. *Object Tracking by Reconstruction with View-Specific Discriminative Correlation Filters*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.

37. Lukezic, A., et al. *Discriminative correlation filter with channel and spatial reliability*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.
38. Bolme, D.S., et al. *Visual object tracking using adaptive correlation filters*. in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2010. IEEE.
39. Bertinetto, L., et al. *Fully-convolutional siamese networks for object tracking*. in *European conference on computer vision*. 2016. Springer.
40. Li, D., et al., *Learning target-aware correlation filters for visual tracking.* Journal of Visual Communication and Image Representation, 2019. **58**: p. 149-159.
41. Sun, Y., et al. *ROI Pooled Correlation Filters for Visual Tracking*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019.
42. Dalal, N. and B. Triggs. *Histograms of oriented gradients for human detection*. 2005.
43. Lowe, D.G., *Distinctive image features from scale-invariant keypoints.* International journal of computer vision, 2004. **60**(2): p. 91-110.
44. Simonyan, K. and A. Zisserman, *Very deep convolutional networks for large-scale image recognition.* arXiv preprint arXiv:1409.1556, 2014.
45. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
46. Zhang, T., C. Xu, and M.-H. Yang, *Learning multi-task correlation particle filters for visual tracking.* IEEE transactions on pattern analysis and machine intelligence, 2018. **41**(2): p. 365-378.
47. Grabner, H., M. Grabner, and H. Bischof. *Real-time tracking via on-line boosting*. in *Bmvc*. 2006.
48. Avidan, S., *Ensemble tracking.* IEEE transactions on pattern analysis and machine intelligence, 2007. **29**(2): p. 261-271.
49. Danelljan, M., et al. *Adaptive color attributes for real-time visual tracking*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014.
50. Zuo, W., et al., *Learning support correlation filters for visual tracking.* IEEE transactions on pattern analysis and machine intelligence, 2018. **41**(5): p. 1158-1172.
51. Sevilla-Lara, L. and E. Learned-Miller. *Distribution fields for tracking*. in *2012 IEEE Conference on computer vision and pattern recognition*. 2012. IEEE.
52. Felsberg, M. *Enhanced distribution field tracking using channel representations*. in *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2013.
53. Ma, C., et al., *Adaptive correlation filters with long-term and short-term memory for object tracking.* International Journal of Computer Vision, 2018. **126**(8): p. 771-796.
54. Taylor, L.E., M. Mirdanies, and R.P. Saputra, *Optimized Object Tracking Technique Using Kalman Filter.* Mechatronics, Electrical Power & Vehicular Technology, 2016. **7**(1).
55. Patel, H.A. and D.G. Thakore, *Moving object tracking using kalman filter.* International Journal of Computer Science and Mobile Computing, 2013. **2**(4): p. 326-332.

56. Tah, A., et al., *Moving object detection and segmentation using background subtraction by kalman filter.* Indian Journal of Science and Technology, 2017. **10**(19).
57. Senna, P., I.N. Drummond, and G.S. Bastos. *Real-time ensemble-based tracker with kalman filter*. in *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. 2017. IEEE.
58. Bailer, C., A. Pagani, and D. Stricker. *A superior tracking approach: Building a strong tracker through fusion*. in *European Conference on Computer Vision*. 2014. Springer.
59. Akram, T., et al., *A deep heterogeneous feature fusion approach for automatic land-use classification.* Information Sciences, 2018. **467**: p. 199-218.
60. Huang, J., et al. *Speed/accuracy trade-offs for modern convolutional object detectors*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
61. Ma, C., et al. *Hierarchical convolutional features for visual tracking*. in *Proceedings of the IEEE international conference on computer vision*. 2015.
62. Xie, D., L. Zhang, and L. Bai, *Deep learning in visual computing and signal processing.* Applied Computational Intelligence and Soft Computing, 2017. **2017**.
63. Wu, Y., J. Lim, and M.-H. Yang. *Online object tracking: A benchmark*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2013.
64. Zhang, J., S. Ma, and S. Sclaroff. *MEEM: robust tracking via multiple experts using entropy minimization*. in *European conference on computer vision*. 2014. Springer.
65. Hong, Z., et al. *Multi-store tracker (muster): A cognitive psychology inspired approach to object tracking*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.
66. Lukežič, A., et al., *Now you see me: evaluating performance in long-term visual tracking.* arXiv preprint arXiv:1804.07056, 2018.
67. Dinh, T.B., N. Vo, and G. Medioni. *Context tracker: Exploring supporters and distracters in unconstrained environments*. in *CVPR 2011*. 2011. IEEE.
68. Kristan, M., et al., *The visual object tracking vot2014 challenge results," 2014*. 2014, C.
69. Shantaiya, S., K. Verma, and K. Mehta, *Multiple object tracking using Kalman filter and optical flow*. European Journal of Advances in Engineering and Technology, 2015. **2**(2): p. 34-39.
70. Liu, C., P. Shui, and S. Li, *Unscented extended Kalman filter for target tracking.* Journal of Systems Engineering and Electronics, 2011. **22**(2): p. 188-192.