VIDEO BEHAVIOR PROFILING FOR ANOMALY DETECTION

THESIS

Submitted in fulfillment of the requirement of the degree of

DOCTOR OF PHILOSOPHY

to

Faculty of Engineering and Technology

by

Neetu Gupta
(Regn. No:YMCAUST-PH18-2011)

Under the Supervision of

Dr. Munish Vashishath

and

Dr. Rajeev Kapoor

Department of Electronics Engineering
J.C. Bose University of Science and Technology, YMCA,
Faculty-121006, (Haryana), INDIA
August 2022

DECLARATION

I hereby declare that this thesis entitled "**Video Behavior Profiling For Anomaly Detection**", by **Neetu Gupta**, being submitted in fulfillment of the requirements for the Degree of Doctor of Philosophy in **Electronics Engineering** under Faculty of Engineering & Technology of J.C. Bose University of Science and Technology, YMCA, Faridabad, during the academic year 2022-23 is a bonafide record of my original work carried out under guidance and supervision of **Dr. Munish Vashishath**, Professor, Electronics Engineering, J. C. Bose University of Science & Technology, YMCA, Faridabad and **Dr. Rajeev Kapoor**, Professor. Electronics and Communication Engineering, Delhi Technological University, Delhi and has not been presented elsewhere.

I further declare that the thesis does not contain part of any work which has been submitted for the award of any degree, either in this university or in any other university.

.

Dated: **Neetu Gupta**

Place: Faridabad **YMCAUST/PH18/2011**

CERTIFICATE

This is to certify that the thesis entitled, "**Video Behavior Profiling For Anomaly Detection**" by **Neetu Gupta**, being submitted in fulfillment of the requirement for the degree of **Doctor of Philosophy** in Electronics Engineering under the Faculty of Engineering & Technology, **J. C. Bose University of Science & Technology, YMCA Faridabad**, during the academic year 2022-23, is a bonafide record of work carried out under my guidance and supervision.

We further declare that to the best of our knowledge, the thesis does not contain any part of any work which has been submitted for the award of any degree, either in this university or in any other university.

Dr. Munish Vashishath
(Professor)
Department of Electronics
Engineering,
J.C. Bose University of
Science & Technology, YMCA
Faridabad, (Haryana)

Dr. Rajiv Kapoor
(Professor)
Department of Electronics and
Communication Engineering,
Delhi Technical University,
Delhi

Date:

ACKNOWLEDGEMENTS

I am deeply indebted to my supervisors **Dr. Munish Vashishath**, Professor, Department of Electronics Engineering, J. C. Bose University of Science & Technology, YMCA, Faridabad and **Dr. Rajiv Kapoor**, Professor, Department of Electronics Engineering, Delhi Technological University, Delhi for their continuous encouragement, invaluable guidance, whole hearted cooperation, moral support, incessant encouragement and appreciation which culminated in the successful completion of this research work. I consider myself extremely fortunate for having got the opportunity to work and learn under their able supervision over the entire period of my association with them. My supervisors with their sharp and inclusive intellect, maestro ability, combined with astute research methodology and deep insight of the subject has unerringly steered the work on smooth and steady course. I wish to express my deepest gratitude to both of them.

I gratefully acknowledge the best wishes and prayers of all my friends for their encouraging, caring words, constructive criticism and invaluable suggestions.

I also express my sincere thanks to Dr. Pradeep Kumar, Professor and Chairperson of Electronics Engineering Department, J.C. Bose University of Science and Technology, YMCA, Faridabad for providing all necessary research facilities in the department. I am also thankful to my colleagues for their support and encouragement during this research work.

I extend my love and gratitude to my two shining stars, my daughter Vaanya and son Purvansh, whose smile was a constant source of inspiration for me.

I dedicate this thesis to my father Mr. Manmohan Agrawal and mother Mrs. Beena Agrawal, who had put their faith in me and found me worthy of achieving this goal.

My special thanks to my sister Dr. Shikha who has been my guiding light all throughout my life. Lastly, my gratitude to the God Almighty whose divine light and warmth provided me the perseverance, guidance, inspiration and strength to complete the work. .

Neetu Gupta
YMCAUST/PH18/2011

ABSTRACT

To ensure the safety of people in public as well as domestic places, surveillance cameras are being installed everywhere such as in banks, malls, markets, academic institutions, parking and most importantly on the roads where the traffic, both pedestrians as well as on wheels, is present 24×7. Different situations are captured by surveillance cameras installed at different places. Accidents on the roads are captured by cameras installed on the roads or traffic lights; crimes and illegal activities are captured by cameras installed in the household, streets, colonies, in hotels and public places. Usually, normal activities happen in daily life but occurrence of some strange event is rare. In order to investigate that event, an intelligent computer vision algorithm needs to be developed which can reduce our time and labor by automatically detecting the strange events occurring at that time.

Most of the existing techniques are either supervised i.e., they require training dataset based on which the test data can be labeled as normal or anomalous. Considering the vast applications of surveillance cameras and the enormous amount of data collected by them, it seems almost impossible to label the entire database as normal or abnormal. A practical Anomaly Detection (AD) system has been developed that runs in real time so as to timely signal an activity that deviates from the normal pattern.

In this research work, object tracking has been proposed for detecting abnormal behavior in videos. By continuously tracking and monitoring an object, a training sample has been developed which is indicative of the normal activity of a person or object of interest. Any deviation from this normal or routine pattern will be considered an anomaly.

This goal of object tracking for Anomaly Detection has been achieved by making use of filters. The filters that have been used in this research work for tracking and anomaly detection are the Bayes Filters. They work on the concept of prediction. The idea is to get samples of the posterior distribution of the hidden states, but since there is no explicit representation of the posterior to draw points from, the prior distribution is sampled, given the observation variables. Thus, it can be said that the posterior probability is based on our prior belief and our observations.

Three kinds of filters have been demonstrated for tracking objects in videos. The first approach uses the Kalman Filter for tracking moving objects. The Kalman Filters show excellent results in Linear Gaussian System but fails to perform in non-linear non-Gaussian environment.

In the second approach particle filters have been used for tracking moving objects. The Particle Filters are suitable for object tracking in non-Gaussian environments with dynamic background thereby outperforming the conventional Kalman Filters but they show poor tracking results in crowded scene where the object of interest might be oc-

cluded by another object.

The third approach proposes the novel Branching Particle Filters that removes the limitations of particle filters. These filters are best suited for tracking and anomaly detection in crowded video datasets in which the object of interest may be occluded by another object.

The validation of the research work is accomplished by comparing the three filters namely Kalman Filter, Particle Filters and Branching Particle Filters on various parameters and also by comparing the proposed technique with previous literature. The comparison concludes that the proposed Branching Filter provides the greatest accuracy in tracking the moving object and thereby detects anomalies in videos better as compared to the other two filters. The computational time and number of iterations taken by it is also less. Findings from this thesis contribute to improve the performance of automated video processing system thereby improving security in areas under surveillance.

CONTENTS

DECLARATION **i**

CERTIFICATE **ii**

ACKNOWLEDGEMENTS **iii**

ABSTRACT **iv**

LIST OF FIGURES **viii**

LIST OF TABLES **x**

LIST OF ABBREVIATIONS **xi**

1 INTRODUCTION **1**

1.1 PREAMBLE 1

1.2 VIDEO BEHAVIOR PROFILING FOR ANOMALY DETECTION . . 3

2 LITERATURE REVIEW **5**

2.1 INTRODUCTION 5

2.2 MOVING OBJECT DETECTION TECHNIQUES 5

2.3 FEATURE EXTRACTION TECHNIQUES 11

2.4 VIDEO BEHAVIOR PROFILING TECHNIQUES 17

2.5 ANOMALY DETECTION TECHNIQUES 26

2.6 MOTIVATION OF RESEARCH 33

2.7 OBJECTIVES 33

2.8 CHAPTER WISE ORGANISATION OF THESIS 34

2.9 SUMMARY 36

3 MODIFIED KALMAN FILTER USING GLOBAL AND LOCAL ORIENTED GABOR TEXTURE HISTOGRAM **37**

3.1 INTRODUCTION 37

3.2 PROPOSED METHODOLOGY 38

3.2.1 Foreground Extraction or Moving Object Detection 39

3.2.2 Object Tracking using Modified Kalman Filter 50
3.3 SIMULATION RESULT 54
3.4 RESULTS AND DISCUSSION 58
3.5 CONCLUSIONS 59

4 MODIFIED PARTICLE FILTER USING GLOBAL AND LOCAL ORIENTED GABOR TEXTURE HISTOGRAM 61
4.1 INTRODUCTION 61
4.2 PROPOSED TECHNIQUE 62
4.2.1 Feature Point Extraction or Feature Detection 63
4.2.2 Feature Description 64
4.2.3 Object Tracking 69
4.3 SIMULATION RESULTS 73
4.4 RESULTS AND DISCUSSION 77
4.5 CONCLUSIONS 78

5 BRANCHING PARTICLE FILTER BASED ANOMALY DETECTION 79
5.1 INTRODUCTION 79
5.2 METHODOLOGY 80
5.2.1 Feature Detection using Harris Detector 81
5.2.2 Feature Description using Optical Flow 81
5.2.3 Tracking using Branching Particle Filter 86
5.3 SIMULATION RESULTS OF BPF TRACKING 88
5.4 ANOMALY DETECTION USING BRANCHING PARTICLE FILTER 93
5.4.1 Motion Descriptor 94
5.4.2 Cluster Motion Patterns 95
5.4.3 Measurement of Similarity 95
5.5 SIMULATION RESULTS OF BPF ANOMALY DETECTION 95
5.6 COMPARATIVE ANALYSIS OF BRANCHING FILTER 98
5.7 RESULTS AND DISCUSSION 103
5.8 CONCLUSIONS 104

6 CONCLUSIONS AND FUTURE SCOPE 105
6.1 CONCLUSIONS 105
6.2 FUTURE SCOPE 108

REFERENCES 111

LIST OF PUBLICATIONS 127

BRIEF PROFILE OF RESEARCH SCHOLAR 129

LIST OF FIGURES

1.1 Generalized Framework for Video Behavior Profiling 3

3.1 Work Flow for Moving Object Detection and Tracking 38
3.2 Frame Difference Method 40
3.3 Image Showing Moving Object 41
3.4 False Detection in Dynamic Background 42
3.5 Block Diagram for Feature Extraction and Kalman Filtering 43
3.6 Depiction of Flat region, Edge and Corner 43
3.7 Choice of Corner Points Using Eigen Values 46
3.8 Harris Corner Detector Applied to the Image of an Athlete 47
3.9 Harris Corner Detector for Different Thresholds 48
3.10 Target Identification by applying GLOGTH on Dataset 1 50
3.11 Successful Tracking of the Target on Applying Kalman Filter (Dataset 1) 54
3.12 Comparative Analysis of Modified Kalman Filter 57
3.13 Poor Tracking Result in Nonlinear environment (Dataset 2) 58

4.1 Anomaly Detection using Particle Filter 62
4.2 Pixel Matrix Generation 67
4.3 Total Gradient Magnitude 67
4.4 Histogram Generation 68
4.5 Tracking Results of Kalman and Particle Filter 74
4.6 Comparative Analysis of Modified Particle Filter 75
4.7 Original video and tracking results of modified particle filter 76

5.1 Anomaly Detection in Videos 79
5.2 Tracking and Anomaly Detection using BPF 81
5.3 Concept of Optical Flow [1] 82
5.4 KTH Dataset, Simple Walk 88
5.5 Tracking Performed on KTH Simple Walk Using BPF 89
5.6 KTH Dataset, Complex Walk 89
5.7 Tracking Performed on KTH Complex Walk using BPF 90
5.8 UMN Dataset, Outdoor Scene 90
5.9 Tracking Performed on UMN Outdoor Scene using BPF 91

5.10 Comparison of Detection Rate of Modified Kalman Filter, Modified Particle Filter, and Branching Filter . . . 91
5.11 Comparison of Accuracy of Modified Kalman Filter, Modified Particle Filter, and Branching Particle Filter . . . 92
5.12 Comparison of Root Mean Square Error of Modified Kalman Filter, Modified Particle Filter, and Branching Particle Filter . . . 92
5.13 Anomaly Detection on UMN dataset using Branching Particle Filter . . 94
5.14 Simulation Results of Anomaly Detection using BPF . . . 97
5.15 Precision vs. Number of Particles . . . 99
5.16 Recall vs. Number of Particles . . . 99
5.17 F Score Curve . . . 100
5.18 ROC and AUROC . . . 101
5.19 RMSE vs. Number of Iterations . . . 101
5.20 Comparative Analysis of Branching Particle Filter in terms of Accuracy and AUC . . . 102
5.21 Comparative Analysis of Branching Particle Filter in terms of Precision, Recall, and F Score . . . 102

LIST OF TABLES

3.1 Performance Comparison of GLOGTH Kalman with Hungarian Kalman on Dataset1 57

4.1 Requirements of a Corner Detector 63
4.2 Performance Comparison of GLOGTH Particle Filter with Hungarian Kalman on Dataset 2 74

5.1 Comparative Analysis of Kalman Filter, Particle Filter & Branching Filter 93
5.2 Performance Metrics for Evaluating the Performance of BPF 98
5.3 Performance Comparison of Proposed work with Previous Literature [2, 3] for UMN Dataset 102

LIST OF ABBREVIATIONS

SYMBOLS	MEANING
AD	ANOMALY DETECTION
ASIC	APPLICATION-SPECIFIC INTEGRATED CIRCUIT
ATM	AUTOMATED TELLER MACHINE
AUROC	AREA UNDER ROC CURVE
BPF	BRANCHING PARTICLE FILTERS
CCA	CANONICAL CORRELATION ANALYSIS
CCTV	CLOSED CIRCUIT TELEVISION
CD-MCBoost	COORDINATE DESCENT MULTICLASS BOOSTING
CFOT	cONSENSUS FOREGROUND OBJECT TEMPLATE
CGS	COARSE GRAIN QUALITY SCALABILITY
CNN	CONVOLUTIONAL NEURAL NETWORK
CRD	CURVILINEAR REGION DETECTOR
CRF	CONDITIONAL RANDOM FIELD
CS	COMPRESSED SENSING
DCT	DISCRETE COSINE TRANSFORM
DNMD	DOUBLE NUCLEAR NORM BASED MATRIX DECOMPOSITION
DoG	DIFFERENCE OF GAUSSIANS
DPCCA	DYNAMIC PROBABILISTIC CANONICAL CORRELATION ANALYSIS
DPCTW	DPCCA WITH TIME WARPINGS
DR	DETECTION RATE
DTN	DELAY/DISRUPTION TOLERANT NETWORKS
FAST	FEATURES FROM ACCELERATED SEGMENT TEST
FN	FALSE NEGATIVE
FP	FALSE POSITIVE
FS	FRAME SEQUENCE
GD-MCBoost	GRADIENT DESCENT MULTICLASS BOOSTING
GLOGTH	GLOBAL AND LOCAL ORIENTED GABOR TEXTURE HISTOGRAM
GMM	GAUSSIAN MIXTURE MODELS
GMT	GENERIC MULTIVIEW TRACKING
GUI	GRAPHICAL USER INTERFACE
HMM	HIDDEN MARKOV MODEL
HOF	HISTOGRAM OF OPTICAL FLOW
HOG	HISTOGRAM OF ORIENTED GRADIENTS
H.O.T	HIGHER ORDER TERMS
HCSD	HUMAN COLOR STRUCTURE DESCRIPTOR
HLCR	HIERARCHICAL LOCALISED CLASSIFICATION OF REGIONS

HSV	HUE SATURATION VALUE
JPEG	JOINT PHOTOGRAPHIC EXPERTS GROUP
KF	KALMAN FILTER
KLD	KULLBACK-LEIBLER DIVERGENCE
KLT	KANADE-LUCAS-TOMASI
KTH	KUNGLIGA TEKNISKA HOGSKOLAN
LBP	LOCAL BINARY PATTERN
LOF	LOCAL OUTLIER FACTOR
LSM	LEAST SQUARES METHOD
MAD	MEAN ABSOLUTE DIFFERENCE
MATLAB	MATRIX LABORATORY
MCMC	MARKOV CHAIN MONTE CARLO
MFCC	MEL FREQUENCY CEPSTRAL COEFFICENTS
MHT	MULTIPLE HYPOTHESIS TRACKING
MoG	MIXTURE OF GAUSSIANS
MOV	EXTENSION FOR MOVIE FILES
MPEG	MOVING PICTURE EXPERTS GROUP
MRF	MARKOV RANDOM FIELD
MSE	MEAN SQUARE ERROR
MSER	MAXIMALLY STABLE EXTREMAL REGIONS
MVC	MULTIVIEW VIDEO CODING
NN	NEAREST NEIGHBOR
ORB	ORIENTED FAST AND ROTATED BRIEF
OT	OBJECT TRACKING
PDF	PORTABLE DOCUMENT FORMAT
PF	PARTICLE FILTER
PPKPF	PROBABILITY PRODUCT KERNEL USING PARTICLE FILTER
QP	QUANTIZATION PARAMETER
RADAR	RADIO DETECTION AND RANGING
RANSAC	RANDOM SAMPLE CONSENSUS
R-CNN	REGIONS WITH CONVOLUTIONAL NEURAL NETWORKS
RFID	RADIO FREQUENCY IDENTIFICATION
RGB	RED GREEN BLUE
RMSE	ROOT MEAN SQUARE ERROR
ROC	RECEIVER OPERATOR CURVE
RPCA	ROBUST PRINCIPAL COMPONENT ANALYSIS
RSS	RECIEVED SINAL STRENGTH
SIFT	SCALE INVARIANT FEATURE TRANSFORM
SIMD	SINGLE STRUCTURE MULTIPLE DATA

SIR	SEQUENTIAL IMPORTANCE RESAMPLING
SMC	SEQUENTIAL MONTE CARLO
SURF	SPEEDED UP ROBUST FEATURE
SUSAN	SMALLEST UNIVALUE SEGMENT ASSIMILATING NUCLEUS
SVM	SUPPORT VECTOR MACHINE
TN	TRUE NEGATIVE
TP	TRUE POSITIVE
UMN	UNIVERSITY OF MINNESOTA
US	ULTRASOUND
VP	VIDEO PROFILING
VS	VIDEO SURVEILLANCE
WEBCAM	WEB CAMERA
WMSN	WIRELESS MULTIMEDIA SENSOR NETWORK

Chapter 1

INTRODUCTION

1.1 PREAMBLE

There is an exponential rise in the demand of automatic methods for analyzing the enormous quantities of surveillance video [1, 2, 3] data generated perennially by closed-circuit television (CCTV) systems. One of the main objectives of deploying an automated visual surveillance system is to detect anomalous behavior patterns and recognize the normal ones [4, 5]. In order to achieve this objective, previously observed behavior patterns need to be studied and analyzed, upon which a criterion is built to decide upon what can be considered as normal and abnormal and then this criteria can be applied to newly captured patterns for detection of anomalies. Due to the vast amount of video surveillance data to be analyzed and the run-time nature of many surveillance applications, it is highly desirable to have an automated system that runs in real time and requires very little or no human intervention. In this research work, it is aimed to develop such a behavior profiling system [6] that is fully unsupervised and presents robust online anomaly detection [7].

The problem of automatic behavior profiling for anomaly detection needs to be identified first. Given a video or an online CCTV input that has been recorded continuously 24×7, the goal of automatic behavior profiling is to develop a model that is capable of detecting abnormal behavior patterns while recognizing instances of expected normal behavior patterns. One of the main challenges for the model is to differentiate anomaly from outliers caused by noisy visual features used for behavior representation. The effectiveness of a behavior profiling algorithm shall be measured by 1) how well anomalies can be detected and 2) how accurately and robustly different classes of normal behavior patterns can be recognized. In this context, an anomaly can be defined as an unusual behavior pattern that is not represented by sufficient samples in a training data set but critically satisfies the constraint to an abnormal pattern. Anything that deviates from a specific routine or path is called anomaly [8].

Video is a combination of large number of frames or images which are arranged in time such that they appear to be continuous to human eye. Video processing is a particular field of signal processing that uses hardware, software or combination of the two for editing the images recorded in video files. VP might include contrast enhancement, noise elimination, pixel size conversion, changing color scheme within an image and detecting anomalies.

To search right structure and contents from visual data, one has to perform video content analysis and it comprises of video parsing and representation with content indexing. Anomaly detection in video surveillance (VS) [9] is tough task as one has to deal with many problems, e.g., noise, illumination change and interaction between different events. Multi-dimensional video sequences are captured often under poor illumination and lighting condition. Multiple cameras may have different positions, orientations and zooming factors. From fundamental point of view, techniques in video investigation are inspired by the need to expand machine learning algorithms that can emulate abilities of human visual frameworks. Moving items in video may be thought of as some compact areas with various obvious movements from background. In particular, if area is moving in specific frame, it is thought of as a moving object all throughout the video. For instance, in video, kid walks for some time and stops to swing his hands. It is important to regard his entire body as moving item rather than his hands. Based on moving object detection [10], we can think of movement as significant prompt to recognize it. If pixel has noteworthy diverse clear movement from background, it generally likely goes with moving object [11]. In like manner, movement signs are characterized as contrast between pixel movement and background movement.

Video investigation innovations can be connected to create keen surveillance frameworks that can be helpful in real time threat identification. Specifically, multi scale tracking can be used for improving the performance of a feature tracker by defining search regions in which lost features can be detected again. Use of visual observation incorporate vehicle and pedestrians traffic checking and human action surveillance for strange action location. A typical surveillance application comprises of three structures: detection of moving item, profiling of behavior and anomaly detection [1] with better movement study. Interactive media frameworks can provide surveillance over a wide range, guaranteeing object perceivability over an enormous area by resolving depth ambiguity problem. Strategies that address the issue of handover between cameras in designs with disjoint perspectives, are showing significant progress. Few products utilizing webcams are available for home and institute campus security through remote surveillance. Almost all webcams come with built in Application Specific Integrated Circuit (ASIC) for the purpose of video compression in real time. Anomaly detection (AD) (also outlier detection) is the identification of items, events or observations which do not conform to an expected pattern or other items in a dataset. Anomaly detection in videos may include detecting a moving object in a place where motion is prohibited or detecting an unidentified or unusual object or stampede or violent behavior detection [12].

1.2 VIDEO BEHAVIOR PROFILING FOR ANOMALY DETECTION

The field of analyzing Video Behavior Profiling [13] is very keen and distinct part in the area of Computer Vision and Video Signal Processing. Video Profiling aims to address the problem of modeling video behavior captured in surveillance videos for the applications of online normal behavior recognition and anomaly detection. Various researchers have proposed several algorithms for Human Activity Recognition and Abnormal Activity Detection, since the analysis of Video Behavior Profiling has numerous applications. For this purpose, a framework is developed for automatic behavior profiling and online anomaly sampling/detection without any manual labeling of the training data set. A generalized framework for Video Behavior Profiling is shown in Figure 1.1.

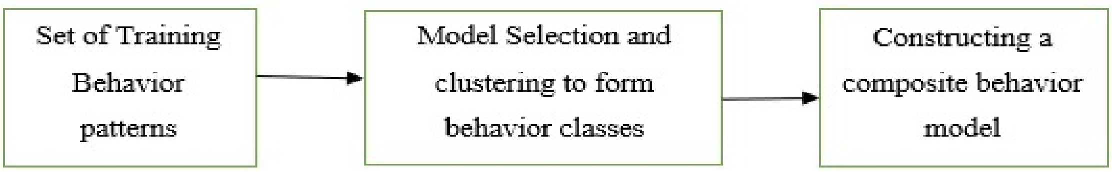


Figure 1.1: Generalized Framework for Video Behavior Profiling

As shown in Figure 1.1 the behavior-based modeling decomposes the behavior space into sub spaces by directly exploring the behavior semantics chalked out by spatio-temporal visual context at the places where the behavior is occurring. Behavior profiling is based on the fact that a large number of objects are influencing each other implicitly or explicitly.

Anomalous activities such as fighting, riots, traffic rule violations and stampede as well as anomalous entities such as weapons at the sensitive place and abandoned luggage should be detected automatically in time. However, the detection of video anomalies is challenging due to the ambiguous nature of the anomaly, various environmental conditions, the complex nature of human behaviors and the lack of proper datasets. Surveillance can help rebuild an incident through the availability of footage for forensics experts thereby helping video analytics. This research work aims to address the problem of profiling video behavior captured in surveillance videos for the purpose of online behavior recognition and anomaly detection.

Chapter 2

LITERATURE REVIEW

2.1 INTRODUCTION

A video surveillance system has the ability to identify the moving objects of video sequence in various computer sight applications as per requirement. Moving object identification methods are used to segment video file into many frames for video surveillance, traffic surveillance, pattern identification and so on. The main objective of Video-object segmentation is to determine and partition the essential objects in video frame from scene background. Video Behavior Profiling [13] is an important job in video surveillance or investigation applications as it provides interrelated temporary data about actuating objects. Video Profiling is the method of describing the actuating object or many objects across time by camera. One of the major essential troubles in computer vision or sight applications is the identification of moving objects. A significant improvement in computer systems has resulted in the introduction of new video object segmentation and detection technique for video surveillance application [14]. In order to determine certain video object detection method, their applications have been extensively developed with the help of literature. The complete literature has been divided into four broad classifications as: Moving Object Detection Techniques, Feature Extraction Techniques, Behavior Profiling Techniques and Video Anomaly Detection Techniques.

2.2 MOVING OBJECT DETECTION TECHNIQUES

The recent technologies for video processing and computer vision have evolved more intelligent since they are required to be aware of video contents like background, object motion and behaviors, scene situation etc. The extraction and analysis of moving objects in a video scene becomes an indispensable function for intelligent systems. The visual information of objects consists of things like- what the objects are, how the objects behave in a footage, how long the objects are present and which direction the objects are headed to.

To extract the object information from a video, the object detection and tracking tool can be utilized in real-time surveillance systems and metadata authoring tools for interactive broadcasting service. In general, the object detection and tracking tool is required to be performed with fast processing speed or in real-time. In case of tracking a criminal in public place, as soon as any object is appeared inside the monitoring screen, the surveillance system has to immediately detect and track that object in real-time.

Izadinia et al. [15] designed a novel method that utilizes correlation between audio-visual dynamics of a video to segment and localize objects that are the dominant source of audio. The approach consists of a two-step spatiotemporal segmentation mechanism that relies on velocity and acceleration of moving objects as visual features. The method can efficiently detect the moving objects associated with sound, whereas it filters out other dynamics in the scene whose motion is uncorrelated to the audio. Moreover, the same framework is exploited for audio-video synchronization as well as interactive video segmentation.

Faro et al. [16] presented a novel vehicle detection and tracking system with stationary camera that relies on a recursive background-modeling approach i.e. the adaptive Poisson Mixture Model which is integrated with a hardware module consisting of luminosity sensors. The luminosity information side channel allows the system to effectively handle rapid changes in illumination, which is one of the typical outdoor applications and bottleneck of the existing Background Pixel Classification Methods.A novel algorithm for detecting and removing partial and full occlusions among blobs was also proposed. The experimental results show that it is a promising technique for the detection and handling of partial occlusions but for handling full occlusions, new algorithms are to be envisaged for sufficient improvements.

A novel algorithm for the real-time unsupervised motion detection in compressed domain sequences was proposed by Patel et al.[17]. The author aimed to develop an algorithm which may be useful in a real-life industrial perspective by facilitating the processing of large numbers of video streams on a single server. The work focussed on using the information in coded video streams to reduce the computational complexity and memory requirements which translates into reduced hardware requirements and costs. The devised algorithm detected and segmented activity based on motion vectors embedded in the video stream without requiring a full decoding and reconstruction of video frames. To improve the robustness to noise, an unreliable motion field was removed by processing motion vector field and Discrete Cosine Transform energy. The algorithm was tested on surveillance H.264 sequences which is a video coding format for recording and distributing video and audio. It provides high accuracy in detecting motion blocks as compared to other algorithms of motion detection but time taken by this algorithm is very high.

A new method of candidate smoke region segmentation is presented by Zhao et al. [18] based on rough set theory. First, Kalman filtering is used to update video background in order to exclude the interference of static smoke-color objects, such as blue sky. Second, in Red Green Blue (RGB) color space, smoke regions are segmented by defining the upper approximation, lower approximation and roughness of smoke-color distribution. Finally, in Hue Saturation Value (HSV) color space, small smoke regions are merged by the definition of equivalence relation so as to distinguish smoke images from heavy fog images. The method lacks in amending the roughness of color distribution to effectively eliminate noise. Moreover, it is not able to adaptively select the threshold which contributes to merging small smoke region.

A method is designed by Xu et al. [19] for motion segmentation by using a moving camera. The method classifies each image pixel in the image sequence as the background or the motion regions by applying a novel three-view constraint called the "Parallax- Based Multiplanar Constraint." The three-view drawback is based on relative projective structure of two points in three different views. The parallax based multi planar constraint addresses the former geometry drawbacks and does not need the reference plane to be constant across multiple views. The work does not reconstruct the static background and the moving objects and does not align them together in the 3D space.

An Adaptive Relighting Algorithm was developed by Guan et al. [20] to improve the brightness uniformity of face images. Facial texture was extracted by using an illumination estimation difference algorithm. Anisotropic Histogram-Stretching Algorithm was proposed to minimize the intraclass distance of facial skin and maximize the dynamic range of facial texture distribution. The method can more effectively eliminate the redundant information of facial skin and illumination.

A robust algorithm is designed by Wu et al. [21] to generate video segment proposals. The proposals generated by the method can start from any frame in the video and are robust to complete occlusions. Though this method does not assume specific motion models and even has a limited capability to generalize across videos, the innovation lies in the use of two efficient moves; the merge move and free addition move. This helps to efficiently start segments from any frame and track them through complete occlusions without much additional computation.

A moving object segmentation algorithm is designed by Wan et al. [22] for freely moving cameras. It is common in outdoor surveillance system, the car build-in surveillance system and the robot navigation system. A two-leveled based affine transformation model optimization method is designed for camera compensation purpose. The outer layer iteration removes the non- background feature points and the inner layer iteration is used to estimate a refined affine model based on the RANSAC method. The feature points are classified into the foreground and background according to the detected motion information. A geodesic based graph cut algorithm extracts the moving foreground depending on classified features.

A novel method named as Hierarchical Localized Classification of Regions (HLCR) was designed by Zhang et al. [23] to cope with difficulties like shape deformation, appearance variations and background clutter. These issues arise in Video Object Segmentation. The method suggests that appearance models as well as the spatial and temporal coherence between frames are the keys to break through bottleneck. In order to identify foreground regions, Hierarchial Localized Classifier was proposed which organizes regional features as decision trees. In global, Gaussian Mixture Color Models (GMMs) was adopted. After integrating the local and global results into a probability mask, the final segmentation result was achieved by graph cut.

Sun et al. [24] aims to tackle the task of semi-supervised video object segmentation across a sequence of frames where only the ground-truth segmentation of the first frame is provided. This approach utilizes reinforcement learning to select optimal adaptation areas for each frame based on the historical segmentation information. This model learns to take optimal actions to adjust the region of interest inferred from the previous frame for online model updating. To speed up the model adaption, a multi-branch tree based exploration method is further developed to fast select the best state action pairs.

An optimized higher-order Conditional Random Field (CRF) is designed by Jiang et al. [25] for automated or machine-driven marking and segmentation of video objects. The approach introduces a computerized optimization scheme to fine tune CRF-associated parameters and make the labeling of segmented regions optimal in formulating the video objects. In comparison with the existing efforts using CRF, optimized

CRF has introduced a number of novel features which can be highlighted as:

1) Higher order CRF labeling is made adaptive to video content changes via a windowed dynamics;

2) Fusion of multiple features is automatically optimized via fuzzy modeling of incoming video content and regression of parameters;

3) Unary potential of higher order CRF labeling is modulated by the shortest path between neighboring regions to improve the effectiveness of higher order CRF labeling and

4) Making the algorithm affordable for a simpler graph-based video segmentation to reduce the overall computing cost, making the algorithm more efficient without compromising on its performances.

A technique for Video Object Segmentation is designed by Ramakanth et al. [26] using patch seams across frames. The technique aims to reduce the image size while preserving the salient image contents. It utilises seams which are connected paths of low energy for retargeting. The energy function associated with the video seams provides temporal linking of patches across frames to accurately segment the object. The energy function takes into account the similarity of patches along the seam, temporal consistency of motion and spatial coherency of seams. Label propagation is achieved with high fidelity in the critical boundary regions, using the patch seams. To achieve this without additional overheads, the error propagation is curtailed by formulating boundary regions as rough-sets.

Zhou et al. [27] proposed a method for learning a distractor-aware discriminative model that can handle continuous missed and inaccurate detection problems due to the occlusion or the motion blur. To deal with target appearance variations, a relational attention learning mechanism is developed to capture the distinctive target appearances by selectively aggregating features from history states with weights extracted from their appearance topological relationship. Based on the discrimination model, a multi-stage tracking pipeline is designed for automatic trajectory initialization, propagation and termination

Yang et al. [28] developed a bottom-up approach for the combination of object segmentation and motion segmentation using a graphical model which is formulated as inference in a Conditional Random Field (CRF) model. This model combines object labeling and trajectory clustering in a unified probabilistic framework. The CRF contains binary variables representing the class labels of image pixels as well as binary variables indicating the correctness of trajectory clustering which integrates dense local interaction and sparse global constraint. An optimization scheme based on a coordinate ascent style procedure is developed to solve the inference problem.

Chien et al. [29] developed a robust segmentation and descriptor based tracking algorithm. Segmentation is applied first and then description for each connected component is extracted for object classification to generate the video object masks. It can do segmentation, tracking, and description extraction with a single algorithm without redundant computation. In addition, a new descriptor for human objects, Human Color Structure Descriptor (HSCD) is also developed for this algorithm. The algorithm can provide precise video object masks and trajectories. HCSD was also able to achieve better performance than scalable color descriptor and color structure descriptor of MPEG-7 for human objects.

Chen et al. [30] proposed to jointly segment and register objects of interest in layered images. The technique is a combination of multiphase active contour method with a joint segmentation registration technique carried out in a local moving window prior to global optimization. To further address layered video sequences and tracking objects in frames, a simple adaptation of optical flow calculations along the active contours in a pair of layered image sequences was developed. The whole integrated algorithm was able to delineate the objects of interest, align them for a pair of layered frames and keep track of the objects over time.

Shin et al. [31] proposed a new framework called Sequential Monte Carlo (SMC) combined with shadow-fading estimation for accurate and attack-tolerant tracking for small-scale mobile primary users. The key idea underlying the technique is to exploit the temporal shadow fading correlation in sensing results induced by the primary user's mobility. The algorithm augments conventional Sequential Monte Carlo (SMC)-based target tracking with shadow-fading estimation. By examining the shadow-fading gain between the primary transmitter and sensors, the algorithm significantly improves the accuracy of primary tracking regardless of the presence or absence of attack. It also successfully masks the abnormal sensing reports due to sensor faults or attacks, preserving localization accuracy and improving spatial spectrum efficiency.

Medeiros et al. [32] presented a parallel implementation of a Histogram-Based Particle Filter for object tracking on smart cameras based on SIMD processors. It specifically focused on parallel computation of the particle weights and parallel construction of the feature histograms since these are the major bottlenecks in standard implementations of Histogram-Based Particle Filters. The algorithm can be applied with any histogram-based feature sets like the simple color histograms as well as more complex Histograms of Oriented Gradients (HOG). One of the major limitations of this approach is that the algorithms are not robust to large variations in the appearance of the target. This is because the color-based approach is sensitive to large illumination changes whereas the HOG-based tracker cannot handle large scale variations.

A Fuzzy Filter is presented by Melange et al. [33] for elimination of random impulse noise in digital grayscale image sequences. The filter consists of different noise detection and filtering steps in which the Fuzzy Set Theory is used. This noise detection is based both on spatial and on temporal information and it is used to prevent the filtering of noise free image pixels. The filtering of the detected noisy pixels is finally performed in a motion compensated way.

Cai et al. [34] proposed a multi-stage object detection architecture called Cascade R-CNN, composed of a sequence of detectors. The detectors are trained sequentially using the output of a detector as training set for the next. This resampling progressively improves hypotheses quality, guaranteeing a positive training set of equivalent size for all detectors and minimizing overfitting. The same technique is applied to eliminate quality mismatches between hypotheses and detectors.

2.3 FEATURE EXTRACTION TECHNIQUES

Feature extraction or dimensionality reduction is an essential part of many machine learning applications. The necessity for feature extraction stems from the issues like the curse of dimensionality and the high computational cost of manipulating high-dimensional data.

Feature Extraction aims to reduce the number of features in a dataset by creating new features from the existing ones (and then discarding the original features). These new reduced set of features are then able to summarize most of the information contained in the original set of features. A summarised version of the original features can be then created from a combination of the original set.

A Tracking-Learning-Data technique is designed by Ding et al. [35]. The author proposes to transfer a generic object tracker to a blur-invariant object tracker without deblurring image sequences. Before object tracking, a large set of unlabeled images is used to learn objects' visual prior knowledge which is then transferred to the appearance model of a specific target. During object tracking, online training samples are collected from the tracking results and the context information. Different blur kernels are involved with the training samples to increase the robustness of the appearance model to severe blur and the motion parameters of the object are estimated in Particle Filter framework.

Nawaz et al. [36] proposed three parameter-independent measures for evaluating multi-target video tracking. The measures take into account target-size variations, combine accuracy and cardinality errors, quantify long-term tracking accuracy at different accuracy levels and evaluate identity changes relative to the duration of the track in which they occur.

Zhang et al. [37] designed a tracking method via multiview learning framework by using multiple Support Vector Machines (SVM). The multiview SVMs tracking method is constructed based on multiple views of features and a combination strategy. To realize a comprehensive representation, three different types of features are selected i.e. gray scale value, Histogram of Oriented Gradients (HOG) and local binary pattern (LBP) to train the corresponding SVMs. These features represent the object from the perspectives of description, detection and recognition respectively. In order to realize the combination of SVMs under the multiview learning framework, a novel collaborative strategy with entropy criterion is developed which is acquired by the confidence distribution of the candidate samples. Moreover, to learn the changes in the object and the scenario, an update scheme is presented based on subspace evolution strategy. The new scheme can control the model update adaptively and help to address the occlusion problems .

An extended Markov Chain Monte Carlo (MCMC) technique and an extended Hidden Markov Model (HMM) technique is designed by Sakaino et al. [38] for learning/recognizing multiple moving objects in videos with jittering backgrounds. A Graphical User Interface (GUI) with improved usability is also designed. This algorithm presents a cost reduction method for the MCMC approach by taking moves out of the iteration loop of the Markov chain when different moving objects interact. For stable and robust tracking, an ellipse model with stochastic model parameters is used. Moreover, HMM method integrates several different modules in order to cope with multiple discontinuous trajectories. The proposed GUI offers an auto-allocation module of symbols from images and a hand-drawing module for efficient trajectory learning and for interest trajectory addition.

A decentralized cooperative technique called Pulse Counting for Disruption Tolerant Network (DTN) localization and a probabilistic tracking method called ProbTracking is presented by Li et al. [39]. Pulse Counting evaluates the user walking steps and movement orientations using accelerometer and electronic compass equipped in cellphones. It estimates user location by accumulating the walking segments and improves the estimation accuracy by exploiting the encounters of mobile nodes. To track user movement, the ProbTracking method uses Markov chain to describe movement patterns and determines the most possible user walking trajectories without full record of user locations.

Wang et al. [40] designed a technique to study the hierarchical features for visual object tracking. First, features robust to diverse motion patterns are learned offline from auxiliary video sequences. The hierarchical features are learned via a two-layer convolutional neural network. Embedding the temporal slowness constraint in the stacked architecture makes the learned features robust to complicated motion transformations which is important for visual object tracking. Then, given a target video sequence, a domain adaptation module is developed to adapt the prelearned features online according to the specific target object. The adaptation is conducted in both layers of the deep feature learning module so as to include appearance information of the specific target object. As a result, the learned hierarchical features can be robust to both complicated motion transformations and appearance changes of target objects.

Song et al. [41] designed an online technique for multi target tracking when the targets are in close proximity or frequently interact with each other. In this system, laser and vision are integrated with tracking and online learning to complement each other in one framework. When the targets do not interact with each other, the laser-based independent trackers are employed and the visual information is extracted simultaneously to train some classifiers online for "possible interacting targets". When the targets are in close proximity, the classifiers learned online are used alongside visual information to assist in tracking. Therefore, this mode of cooperation not only deals with various tough problems encountered in tracking but also ensures that the entire process can be completely online and automatic.

Wang et al. [42] designed an algorithm that transfers visual prior learned offline for online object tracking. From a collection of real-world images, an overcomplete dictionary is learned to represent visual prior. The prior knowledge of objects is generic and the training image set does not necessarily contain any observation of the target object. During the tracking process, the learned visual prior is transferred to construct an object representation by sparse coding and multiscale max pooling. With this representation, a linear classifier is learned online to distinguish the target from the background and to account for the target and background appearance variations over time. Tracking is then carried out within a Bayesian Inference Framework, in which the learned classifier is used to construct the observation model and Particle Filter is used to estimate the tracking result sequentially.

An empirical analysis was conducted by Song et al. [43] to explore the basic laws governing human mobility following disasters and an effective human mobility model is developed to predict and simulate population movements.The algorithm suggest that human mobility following disasters can be significantly more predictable and be more easily simulated than previously thought.

Salti et al. [44] designed a unified conceptual framework for appearance model adaptation that enables a principled comparison of different approaches. Moreover, a novel evaluation methodology was introduced that enables simultaneous analysis of tracking accuracy and tracking success without the need of setting application-dependent thresholds. Based on the framework and this evaluation methodology, an extensive experimental comparison of trackers was conducted that perform appearance model adaptation.

Deori et al. [45] proposed a literature review on the state of the art tracking methods, categorized them into different categories and then identified useful tracking methods. Most of the methods included object segmentation using background subtraction. The tracking strategies used different methodologies like Mean-Shift, Kalman Filter, Particle Filter etc. The performance of the tracking methods varied with respect to background information. In this survey, the author discussed the feature descriptors that are used in tracking to describe the appearance of objects which are being tracked as well as object detection techniques.

Li et al. [46] provides a detailed review of the existing 2D appearance models. In particular, this survey takes a module-based architecture to easily grasp the key points of visual object tracking. In this survey, first the problem of appearance modeling is decomposed into two different processing stages namely, visual representation and statistical modeling. Then, different 2D appearance models are categorized and discussed with respect to their composition modules. Finally, several issues of interest are addressed as well as the remaining challenges for future research are discussed .

Jadhav et al. [47] presented an approach of object detection for video surveillance. The algorithm consists of various steps including video compression, object detection, and object localization. In video compression, the input video frames are compressed with the help of two-dimensional Discrete Cosine Transform (2D DCT) to achieve less storage requirements. In object detection, key feature points are detected by computing the statistical correlation and the matching feature points are classified into foreground and background based on the Bayesian Rule. Finally, the foreground feature points are localized in successive video frames by embedding the maximum likelihood feature points over the input video frames.

Sukanyathara et al. [48] proposed a framework which can effectively detect the moving objects and track them despite of occlusion and a priori knowledge of objects in the scene. The segmentation step uses a robust threshold decision algorithm which uses a multi-background model. The video object tracking is able to track multiple objects along with their trajectories based on Continuous Energy Minimization. In this work, an effective formulation of multi-target tracking as minimization of a continuous energy is combined with multibackground registration. It focuses on making use of an energy that corresponds to a more complete representation of the problem rather than one that is amenable to global optimization. Besides the image evidence, the energy function considers physical constraints such as target dynamics, mutual exclusion and track persistence. The tracking framework is able to track multiple objects despite of occlusions under dynamic background conditions.

A high-capacity steganography scheme is proposed by Zhang et al. [49] for the Joint Photographic Expert group (JPEG 2000) baseline system which uses bit-plane encoding procedure twice to solve the problem due to bitstream truncation. Moreover, embedding points and their intensity are determined in a well defined quantitative manner via redundancy evaluation to increase hiding capacity. The redundancy is measured by bit which is different from conventional methods which adjust the embedding intensity by multiplying a visual masking factor. High volumetric data is embedded into bit-planes as low as possible to keep message integrity but at the cost of an extra bit-plane encoding procedure and slightly changed compression ratio. The method can be easily integrated into the JPEG 2000 image coder and the produced stegnographed bitstream can be decoded normally.

Bhattacharya et al. [50] discussed the challenges of automating surveillance and reconnaissance tasks for infra-red visual data obtained from aerial platforms. The author gave a study of various image registration techniques that are required to eliminate motion induced by the motion of the aerial sensor. Next, a technique for detecting moving objects from the ego-motion compensated input sequence was presented. Finally, a methodology for tracking already detected objects is described using their motion history.

Sun et al. [51] presented an automatic foreground object detection method for videos captured by freely moving cameras. The approach does not require any training data nor the interaction by the users. Furthermore, a probabilistic Consensus Foreground Object Template (CFOT) is constructed which is directly applied to the input video for moving object detection via template matching. CFOT can be used to detect the foreground object in videos captured by a fast moving camera even if the contrast between the foreground and background regions is low. The method can be generalized to foreground object detection in dynamic backgrounds and is robust to viewpoint changes across video frames.

Han et al. [52] designed a ”coarse to fine” method to avoid raster scanning entire image. Foreground pixels are detected in coarse level to roughly locate the foreground objects in the image and then finer detection is performed on the corresponding blocks gradually. Secondly, fast mean shift approach is presented according to temporal dependencies. Mean shift iterations are performed starting from incoming data and the mode found in last time.

Vahora et al. [53] described various approaches of moving object detection such as background subtraction, temporal difference as well as pros and cons of these techniques. The presented method tries to overcome the problem of superfluous effects of foreground objects as well as reduces the computational complexity up to some extent. A robust algorithm for automatic, noise detection and removal from moving objects in video sequences is presented. The algorithm considers static camera parameters.

Tong et al. [54] designed a new metric for video encryption which evaluates visual security based on color and edge features of original and cipher-videos. The metric is easy to be incorporated into video encryption system for visual security based encryption decision. In addition, subjective tests for visual security assessment have been fully carried out.

Zhang et al. [55] presented a method called Double Nuclear Norm-based Matrix Decomposition (DNMD) for dealing with the image data corrupted by continuous occlusion. The method uses a unified low-rank assumption to characterize the real image data and continuous occlusion. Compared with Robust Principal Component Analysis (RPCA), the low-rank assumption of DNMD is more intuitive for describing occlusion. Moreover, DNMD is solved by alternating direction method of multipliers.

A compressive video sensing encoder is planned by Pudlewski et al. [56]. This model is used to reduce the required energy and computational complexity at the source node. The encoder leverages the properties of Compressed Sensing (CS) to overcome many of the limitations of traditional encoding techniques specifically lack of resilience to channel errors, and high computational complexity. An analytical/empirical model based on the notion of rate-energy-distortion is developed. It predicts the received video quality when the overall energy available for both encoding and transmission of each frame of a video is fixed and limited and the transmissions are affected by channel errors.

2.4 VIDEO BEHAVIOR PROFILING TECHNIQUES

Video profiling mainly correspond with moving image, target object images and background which change over time. The video profiling process in video sequence is based on object's representation being employed. Video profiling for anomaly detection in video sequences is a demanding task.

Video Profiling as described by Kothiya et al. [57] is a technique of locating moving objects over time using the camera in video sequences in real time. The objective of video profiling is to associate target objects in consecutive video frames. VP requires location and shape or features of objects in the video frames. Every profiling algorithm needs to detect moving object. Hence, object detection is the preceding step of VP in computer vision applications. After that, detected object can be extracted by the feature of moving object to track that moving object in video scene. It is a challenging task in image processing to track the objects in consecutive frames. Various challenges can arise due to complex object motion, irregular shape of object, occlusion of object by another object and real time processing requirements. VP has a variety of uses, some of which are: surveillance and security, traffic monitoring, video communication, robot vision and animation.

Video Profiling system initially takes video frames as input and then performs preprocessing for removing noise frames in input video. Finally, VP system performs moving object detection and tracking [58]. Recently, number of methods have been designed for VP. Some of the models presented for video profiling are contour and feature point-based models. Contour based profiling is accomplished by locating the object region in every frame with the help of object model that is created by the previous frames. Boundary silhouettes represent object shapes and the tracking results so obtained in the video frames are updated dynamically. Hu et al. in [59] developed a Markov Model based dynamical shape method for periodic motions. However, active contour-based visual tracking framework lacks tracking accuracy.

Region-Based profiling represents object based on color. Therefore, region-based object model is computationally efficient. But, its computationally efficiency is corrupted when some objects move jointly in image sequences. Besides, when multiple objects move simultaneously then due to occlusion, region-based object model lacks in its tracking accuracy. Varas et al. in [60] designed region-dependent Particle Filter for general target tracking which integrates color-feature with Particle Filter. This algorithm significantly performs video profiling and gives precise segmentation of target.

Feature Point Based VP Algorithm is based on calculating the local minimum energy points around a feature point thereby analyzing the features of image tensor. There are three fundamental steps in this algorithm. First step is to identify and track object through extracting elements. Second step is to group them into higher level characteristics. The final step is to evaluate these features in subsequent frames. Weinberger et al. [61] presented new tracking framework that employs Distance Metric Learning in combination with Nearest Neighbor (NN) classification for Object Tracking. Canny Edge Detector using NN classifier was employed in this method for detecting objects. It was able to distinguish object from other objects and subtract background from frame with help of NN algorithm. It employs the technique of calculating the distance between object and background. The Blob Detector is then used to identify object based on skin color. A bounding box is then constructed for recognizing the object.

Profiling in videos finds immense applications in the areas of wildlife monitoring, traffic control etc. A number of techniques are available to serve the purpose. Some techniques need the central step of object segmentation to track the target. It is the process of segmenting object/target from the rest of the video sequence. The object or target may be then classified based on shape, color, texture or any other feature. Next step is to track the detected object. Object tracking (OT) algorithms may be broadly classified into three categories namely Point Based, Kernel Based and Silhoutte Based Tracking.

1. Point-Based OT can be defined as the correspondence of detected objects represented by points across frames. Point correspondence may pose complications-mainly in the case of existence of occlusions, misdetections, entry and exit of objects. Point correspondence methods can be divided into two broad categories namely deterministic and statistical methods [62]. The deterministic methods use qualitative motion heuristics to constrain the correspondence problem. On the other hand, probabilistic methods explicitly take the object measurement and take uncertainties into account to establish correspondence.

    (a) Deterministic Method works on connecting each object in previous frame with single object in current frame. This is done with the help of set of motion constraints. Deterministic methods for point correspondence define a cost of associating each object in previous frame to a single object in current frame using a set of motion constraints. Minimization of the correspondence cost is formulated as a combinational optimization problem.

(b) Statistical or Probabilistic Method works on measuring position of object in the frame with detection mechanism. This method is used for modeling the object properties such as velocity and position. Measurements obtained from video sensors invariably contain noise. Moreover, the object motions can undergo random perturbations e.g. maneuvering vehicles. Statistical correspondence methods solve these tracking problems by taking the measurement and the model uncertainties into account during object state estimation. The state space approach is used by statistical correspondence methods to model the object properties like position, velocity and acceleration. Measurements usually consist of the object position in the image which is obtained by a detection mechanism. Thus, statistical method is an extended endeavor of deterministic method by using different kinds of filters. The commonly used filters are Multiple Hypothesis Filters, Kalman filter, Particle filter and Joint Probability Data Association Filter.

i. Multiple Hypothesis [63] represents a significant generalization to classical estimation problems which considers an unknown and time-varying number of objects. Sensor measurements are generally collected at different sequence of times and exhibit several challenges that include missed detections, false alarms and even if object measurements are available, challenges like noise-corrupted observation of object states are experienced. A fundamental challenge is measurement-origin uncertainty i.e. it is not known which measurements are associated with the same object and which measurements are to be discarded as false alarms.

ii. Kalman Filter [64] is essentially a set of mathematical equations that implement a predictor-corrector type estimator that is optimal in the sense that it minimizes the estimated error covariance when some presumed conditions are met. The Kalman Filter estimates a process by using a form of feedback control. the Filter estimates the process state at some time and then obtains feedback in the form of noisy measurements. The equations for the Kalman Filter fall into two groups, time update equations and measurement update equations. The time update equations are responsible for projecting forward (in time) the current state and error covariance estimates to obtain the a priori estimates for the next time step. The measurement update equations are responsible for the feedback i.e. for incorporating a new measurement into the a priori estimate to obtain an improved a posteriori estimate.

iii. Particle Filter [65] is the approximation of relevant probability distributions using the concepts of sequential importance sampling and approximation of probability distributions using a set of discrete random samples with associated weights. PF does not involve the linearization approximating to nonlinear systems that is required by the Extended Kalman Filter. PF methods still need to be improved in the aspects of accuracy and calculating speed.

2. Kernel-Based Profiling [66] provides a robust solution to address the most common problems in tracking i.e. representation and localization. In this technique an object is represented by a feature histogram with an isotropic kernel. To localize the kernel, gradient based mean shift optimization is performed. The method has proved to be successful for non-rigid motion containing significant clutter. Though robust, the algorithm may show poor performance under cases of occlusion. Kernel Based Tracking Methods can be further categorized into:

   (a) Simple Template Matching [67] is the best strategy in image processing for finding small parts of an image which coincide or match with the reference image called the template. Template based matching may require sampling of large number of points. To make improvements to the technique, more than one template having different scales and rotations can be used.

   (b) Mean Shift Method [68] is a recently introduced feature space analysis technique which attempts to locate the densest point of samples iteratively. The most glaring advantage of this algorithm is that unlike Gaussian mixture models and k means, it makes no model assumptions. The mean shift algorithm is also able to model complex clusters though the algorithm also holds the disadvantage of breaking down in high dimensions.

   (c) Support Vector Machine (SVM) [69] is a classifier which is used to find a hyperplane. The hyperplane can be considered as a line which divides the space into two parts, one which contains datapoints belonging to one class (may be called as positive) and other containing datapoints belonging to another class (may be called as negative). SVM comes under the category of supervised learning models that analyze data for classification and abnormality detection.

(d) Layering Based VP highlights that major issues arise from large variation in target saliency and significant motion heterogeneity which may result in the failure of tracking weak targets. To tackle this challenge, a hierarchical layered tracking structure is used to perform tracking sequentially layer by layer. Upon this layered structure, an inter target mutual assistance mechanism is established on basis of inter target correlation exploited among targets. The tracking results of a subset of targets can be utilized as additional information for tracking other targets [70].

(e) Multi-View Based VP highlights that recent advancements in visual tracking have immensely improved the tracking performance. However, challenges like occlusion and view change remain obstacles in real world deployment. An easy solution to these challenges is to use multiple cameras with multiview inputs. Another challenge is that the existing systems are mostly limited to specific targets, use static cameras and require camera calibration. To break through these limitations, a Generic Multiview Tracking (GMT) framework [71] is developed, that allows camera movement while requiring neither specific object model nor camera calibration. This type of tracking implicitly and dynamically encodes camera geometric relations and addresses missing target issues such as occlusion.

3. Silhouette Based VP is performed by estimating the region of object of interest in each frame. Silhouette tracking methods [72] use the information encoded inside the object region. This information can be in the form of appearance, density and shape models which are usually in the form of edge maps. Given the object models, silhouettes are tracked by either shape matching or contour evolution.

   (a) Contour Evolution VP [73] is carried out by locating the object region in every frame through the object model created by the previous frames. The object shapes are considered as boundary silhouettes and the tracking results obtained are updated dynamically in the video frames.

(b) Shape Matching process necessitates that specific measure of article in current edge overlay with item district in past casing. Shape Matching [74] can be employed utilizing two distinct methodologies. First part uses state space models to demonstrate shape and movement. Second approach legitimately develops form by limiting shape vitality utilizing direct minimization systems, e.g. inclination plummet.

Kamp et al. [75] designed animated 3D models embedded in PDF documents that combines the benefits of both interactivity and movie. The methods simulates joints with largely deterministic movements due to precise form closure. The function of an individual nut and screw kind hip joint is explained & proceeded to complex movements. The basic posture is attained using particular cascade movements. Legs and Head interlock mutually with particular characteristics of first abdominal ventrite and thorax to improve the stability of beetle and to preserve the defense structure or position with less muscle activity.

Afonso et al. [76] presented an automatic method for building pedestrian trajectories in far field surveillance scenarios not requiring user intervention. This basically comprises the detection of multiple moving objects in a video sequence through the detection of the active regions followed by the estimation of the velocity fields that is accomplished by performing region matching of the above regions at consecutive time instants. This leads to a sequence of centroids and corresponding velocity vectors which describe the local motions presented in the image. A motion correspondence algorithm is applied to group the centroids in a contiguous sequence of frames into trajectories corresponding to each moving object. The method also contributes for automatically finding the trajectories from a library of previously computed ones. Motion fields can be reliably estimated from these automatically detected trajectories leading to a fully automatic procedure for the estimation of multiple motion fields.

A new hierarchical moving target detection technique is designed by Hao et al. [77] depending on spatiotemporal saliency. Temporal saliency is used to obtain a coarse segmentation and spatial saliency is extracted to obtain the object's appearance details in candidate motion regions. By combining temporal and spatial saliency information, refined detection outputs are obtained. For full description of object distribution, spatial saliency is identified in both region and pixel levels depending on the local contrast.

An algorithm is designed by Chiranjeevi et al. [78] for improving object detection under heavy dynamic background conditions by modeling uncertainties in the data by interval-valued fuzzy set. Real-valued fuzzy aggregation was extended to interval-valued fuzzy aggregation by considering uncertainties over real similarity values. The algorithm calculates the uncertainty that varies for each feature, at each pixel and at each time instant. It adaptively determines membership values at each pixel by the Gaussian of uncertainty value instead of fixed membership values thereby giving importance to a feature based on its uncertainty.

A new system is planned by Dopico et al. [79] for search and identification of moving objects in sequence of images collected by camera in conventional vehicle. The key aim is to design and execution of software system depending on optical flow analysis for recognition of the moving objects by driver. The optical flow is computed for all image sequences for calculating the motion. Two segmentation methods are carried out namely, optical flow and the images of sequence. The segmentation method detects the movement of objects in sequence with direction and magnitude.

Multiview Video Coding (MVC) model is designed by Stefania et al. [80] to improve the performance in Wireless Multimedia Sensor Networks (WMSN). The technique improves performance by leveraging the spatial correlation among partially overlapped fields of view of multiple video cameras observing the same scene. The technique concludes, that in addition to geometric information, occlusions and movement need to be considered to fully take advantage of multiview video coding.

A new reconfigurable model is designed by Lin et al. [81] for face detection and alignment. The author introduced the facial landmark localization method based on a cascaded fully convolutional network. This method first generates low-resolution response maps to identify approximate landmark locations and then produces fine-grained response maps over local regions for more accurate landmark localization. It then introduced the attention-aware facial hallucination method which generates a high-resolution facial image from a low-resolution image. This method recurrently discovers facial parts and enhances them by fully exploiting the global interdependency of facial images.

Lee et al. [82] designed a visual recognition system operating on a hand-held device based on a video based feature descriptor that characterizes its invariance and discriminative properties. Feature selection and tracking are performed in real-time and used to train a template-based classifier during a capture phase prompted by the user. During normal operation, the system recognizes objects in the field of view based on their ranking. The technique also exploits the characterization of the stability properties of local invariant detectors and of the conditions under which a template-based descriptor is optimal.

A hybrid method is designed by Chunxian et al. [83] to identify the moving vehicle in aerial videos. Local feature extraction and matching are carried out to calculate the global motion. It is achieved with the Speeded Up Robust Feature (SURF) key points for stabilization and improved performance. A list of dynamic pixels was obtained and collected for many moving vehicles with different optical flows. For improving the detection accuracy, preprocessing techniques are used in the surveillance system like road traffic monitoring etc.

Chen et al. [84] designed computer vision tools to automate the collection and distribution of audio visual content. This technique is a complete production process of personalized video summaries in a typical application scenario where the sensor network for media acquisition is composed of multiple cameras. The process involves numerous integrated technologies and methodologies including automatic scene analysis, camera viewpoint selection, adaptive streaming and generation of summaries through automatic organization of stories. The technology provides practical solutions to a wide range of applications such as personalized access to local sport events through a web portal, cost-effective and fully automated production of content dedicated to small-audience or even automatic login of annotations.

A method is planned by Kermani et al. [85] for recognition of moving objects in surveillance applications. The method combines an adaptive filtering method with Bayesian Change Detection Algorithm. The adaptive structure firstly detects the edges of motion objects. Bayesian Algorithm then corrects the shape of identified targets. The technique exhibits considerable robustness against noise, shadows, illumination changes and repeated motions in the background. In the algorithm, no prior information about foreground and background is required and the motion detection is performed in an adaptive scheme.

Mansour et al. [86] designed a single layer scalable video rate and distortion model for video bitstreams encoded using Coarse Grain Quality Scalability (CGS) feature of the scalable extension. The source is Laplacian distributed and compensate for errors in distribution assumptions by linearly scaling the Laplacian parameter. The simplified approximation of models allows the runtime computation of sequence dependent model constants. The models deploy the Mean Absolute Difference (MAD) of the prediction residual signal and the encoder quantization parameter (QP) as input parameters.

A nonlinear mixing model is designed by Halimi et al. [87] for hyper spectral image unmixing and nonlinearity recognition. Appropriate priors are chosen for its parameters to satisfy the positivity and sum-to-one constraints for the abundances. The joint posterior distribution of the unknown parameter vector is then derived. The developed Metropolis-within-Gibbs algorithm allows samples distributed according to the posterior to be generated and to estimate the unknown model parameters.

Yang et al. [88] designed a system to use spatial information of Received Signal Strength (RSS) inherited from wireless nodes to detect the spoofing attacks. Cluster-based mechanisms are developed to determine the number of attackers. When the training data are available, the Support Vector Machines (SVM) method is deployed to further improve the accuracy of determining the number of attackers. In addition, an integrated detection and localization system is designed that can localize the positions of multiple attackers.

A new framework is planned by McFee et al. [89] for multiple class object localization that incorporates different levels of contextual interactions. The contextual interactions are based on three different sources of context. The framework learns a single similarity metric from multiple kernels, combining pixel and region interactions with appearance features and then applies a conditional random field to incorporate object level interactions. For combining many feature descriptions, large margin nearest neighbor that supports many multiple kernels is deployed.

The problem of localizing specific anatomic structures using ultrasound (US) video is addressed by Kwitt et al. [90] by the increased availability of portable and low-cost US probes. The localization approach is motivated by dynamic texture analysis and leverages many advances in the field of activity identification.

Data hiding techniques are designed by Shanableh et al. [91] by applying compressed MPEG video. The initial technique hides the message bits by changing the quantization scale of a constant bit rate video. A payload of one message bit per macro block is achieved. A second order multivariate regression is generally used to find out the connection of macro block-level feature variables with the values of hidden message bit. The regression model is then used by the decoder to predict the values of the hidden message bits with very high prediction accuracy. The method employs the adjustable macro block ordering characteristics to conceal message bits.

Nakhmani et al. [92] designed an algorithm for contour self-crossing detection. This model is basically designed for parametric active contours. An easy method is designed via simple techniques from differential topology. The detection is achieved by examining the entire net variation of contour's angle without point sorting and plane sweeping. The additional algorithms are planned for locating crossings by angle considerations and by plotting the four-connected lines between the discrete contour points.

Ling et al. [93] designed a learning-based approach to increase the temporal resolutions of human motion sequences. Given a set of high resolution motion sequences, first the motion tendency from this learning dataset is learned and then new postures for the low-resolution sequence are synthesized according to the learned motion tendency. Each motion sequence is first projected into a low-dimension manifold space where the local distance between postures could be better preserved. Then each of the projected motion sequences is represented as a motion trajectory. Next, motion priors learned from the high resolution training sequences are used to reconstruct the motion trajectory for the input sequence. Finally, the reconstructed motion trajectory is combined with object inpainting technique to generate the final result.

Nicolaou et al. [94] planned Dynamic Probabilistic Canonical Correlation Analysis (DPCCA). A generative model determines the temporal dependency on shared or individual spaces. For the temporal lags with continuous annotations, a latent warping process resulting in DPCCA with time warpings is planned (DPCTW). Thus, DPCCA and DPCTW were extended to two supervised variants to include inputs and observations both in a generative and discriminative framework.

2.5 ANOMALY DETECTION TECHNIQUES

Anomaly detection is an important problem that has been encountered in diverse research areas and application domains. Many anomaly detection techniques have been specifically developed for certain application domains, while others are more generic. Anomaly detection is the documentation of rare explanations which raise doubts by contrasting with most of the data. Usually, anomalous things indicate some sort of issue such as bank fraud, therapeutic issues or mistakes in content. Anomalies are additionally called as abnormalities, outliers, noise, deviations and exceptions.

Various anomaly identification methods are available [95] to provide such information unless the data is collected properly. Many data sets continuously stream from weblogs, financial transactions, health records and surveillance logs, as well as from business, telecommunication and biosciences. Many existing anomaly detection techniques fail to retain sufficient accuracy due to the so called "Big Data" characterised by high-volume, and high-velocity data generated by variety of sources. High dimensionality creates difficulties for anomaly detection because when the number of attributes or features increase, the amount of data needed to analyze the content accurately, also increases. This results in data sparsity in which data points are more scattered and isolated. This data sparsity is due to unnecessary variables or the high noise level of multiple irrelevant attributes that conceal the true anomalies. One way to address the problem of anomaly detection in high dimensional big data is to use a triangular model of vertices. Other way to address the problem of high dimensionality is to reduce the dimensionality which projects the whole data set into a lower dimensional space either by combining dimensions into linear combinations of attributes or by selecting the subsets of locally relevant and low-dimensional attributes called subspaces.

Anomalies are divided into three types namely point anomaly, contextual anomaly and collective anomaly. Point anomalies are outliers that exist far outside the rest of the data e.g. in business use case, credit card fraud can be detected based on ”amount spent”. In contextual anomaly, information is anomalous in particular setting or specific context. It may also be termed as restrictive anomaly. Attributes of data objects are divided into two groups namely contextual and behavioral. The attributes that provide only contexts of the behavior are contextual characteristics. The attributes that are usually highly related to the outlier behavior are called behavioral attributes [96]. Collective Anomalies may be defined as a group of instances in which each particular instance of the group may not be anomalous itself but may prove to be anomalous when these instances occur collectively.

Common methods available for finding anomalies are classification based, statistical based and clustering based. Classification based method of anomaly detection uses auxiliary transformations for training a model to extract useful features. Classification based anomaly detection method is categorized into two steps. In the first step, the training phase acquires a model with the marked training data set.In the second step, the testing phase divides a test data in simple form with the help of the model learnt in the first step.

Some classification based [97] techniques are the Support Vector Machines-based Techniques, Neural Networks-based Techniques, Nearest Neighbor-based Techniques etc.

The basic idea of any statistical anomaly method is that an anomaly is an observation which is suspected of being partially or wholly irrelevant because it is not generated by the stochastic model. This method assumes that the normal data instances settle at higher area of a stochastic model. The validity of many statistical procedures relies on the assumption of approximate normality but if the data is not normal then nonparametric procedures might be the probable solution for handling abnormal data. Statistical methods are broadly classified as parametric and nonparametric. A statistical procedure is said to be of nonparametric type if its properties get satisfied by reasonable approximation under assumptions of reasonably moderate nature.

Clustering-Based Detection Technique is generally employed to categorize same data instances and this is basically an unsupervised type method. The first type of Clustering method makes the assumptions that normal data lies in a cluster and that the anomalies do not relate to any datapoint. The second type makes the assumption that normal data instances relate to big and dense cluster and that the hand anomalies may be related to small clusters.

Saberian et al.[98] planned two new multiclass boosting algorithms. The first algorithm called CD-MCBoost is a coordinate descent procedure that updates one predictor component at a time. The second algorithm called GD-MCBoost is a gradient descent procedure that updates all components jointly. The formulation enables a unified treatment of many previous multiclass boosting algorithms.

Ercan et al. [99] planned a task driven method with camera subset. In the task-driven approach, each camera first performs simple local processing to detect the horizontal position of the object in the image. This information is then sent to a cluster head to track the object. The locations of the static occluders are assumed but only prior statistics on the positions of the moving occluders are available. A noisy perspective camera measurement model is introduced where occlusions are captured through occlusion indicator functions. An auxiliary particle filter that incorporates the occluder information is used to track the object. The camera subset selection algorithm uses the minimum mean square error of the best linear estimate of the object position as a metric and tracking is performed using only the selected subset of cameras.

Pan et al.[100] designed a text area detector. This model is designed to calculate the text existing confidence and scale data in image pyramid to partition the text constituents by local binarization. For removing the non text components, a conditional random field (CRF) model with unary or single constituent features and binary contextual constituents' component relationships with supervised parameter learning is planned. Text constituents or components are grouped into text words with a learning-based energy minimization method.

Yang et al. [101] planned contour-based object detection. A solution is provided to the problem by locating the dominant sets in weighted graphs. The nodes of the graph are pairs composed of model contour parts and image edge fragments and the weights between nodes are based on shape similarity. Because of high consistency between correct correspondences, the correct matching corresponds to a dominant set of the graph. When a dominant set is determined, it provides a selection of correct correspondence. The proposed method is able to get all the dominant sets thereby detecting multiple objects in an image in one pass.

Lemaitre et al. [102] planned a method for a wiry target detection and matching based on a new Curvilinear Region Detector (CRD) and a shape context-like descriptor. The standard techniques for local patch detection are not suitable to wiry objects and curvilinear structures like railroads and rivers in satellite images. The detection process is estimated using the segmentation quality of curvilinear regions.

Walha et al. [103] designed a moving object detection system based on camera motion estimation. Local feature extraction and matching is used to estimate global motion and Scale Invariant Feature Transform (SIFT) keypoints are found to be suitable for the stabilization task. Moving object is then detected by Kalman Filtering. Motion compensation is carried out to obtain a stabilized video sequence.

Kang et al. [104] designed a Compressed Sensing (CS) based algorithm for the detection of moving object in video sequences. First, an object detection model is developed to simultaneously reconstruct the foreground, background and video sequence using the sampled measurement. Then, the reconstructed video sequence is used to estimate a confidence map to improve the foreground reconstruction result.

Wang et al. [105] presented a novel 3D spatiotemporal Difference-of-Gaussians (DoG) Filter-based algorithm for tracking small targets in videos of various complex scenes. This model is generally designed for small target tracking which are capable of accounting for spatial and temporal information. Based on such filters, an effective and robust tracker is constructed to track the small targets which are spatiotemporally distinguishable from background clutter.

An efficient technique is designed by Hu et al. [106] to identify the moving objects for videos captured by moving camera. Moving object detection is complex in videos captured using moving camera. This is because in the videos filmed by moving cameras not only do the objects move but also the frames shift. In this algorithm, the feature points in frames are identified and divided into the foreground and background. The foreground regions and image difference are obtained and then merged to obtain moving object contours. The moving object is identified depending on the motion history of continuous motion contours and refinement schemes.

A spatial–temporal sampling process is described by Chuang et al.[107] as unified method of extracting the video objects and computing the spatial–temporal boundaries using a learnt video object model. A basic approach is designed for learning optimal key target/object codebook sequence from a set of training video clips to explain the semantics of identified video objects. The dynamic programming is employed to find the video objects with spatial–temporal boundaries in a test video clip. .

Ferryman et al. [108] planned a video surveillance framework that identifies the desolated objects or target in observation scenes. The scheme is based on threat assessment algorithm that combines the idea of ownership with automatic understanding of social relations in order to infer abandonment of objects. The execution is carried out with the development of logic-based inference engine and the threat detection execution is carried out by testing against a range of datasets describing realistic situations.

Corner based approach is planned by Zhao et al. [109] to identify the text and caption from videos. The approach is encouraged by the observation that there exist corner points in characters especially in text and caption. Many discriminative features are designed to explain the text regions created by corner points. The utilization of these features in flexible way leads to its deployment in many applications. Language independance is an important advantage of this method. Another algorithm is designed to identify moving captions in videos. The motion features extracted by optical flow are combined with text features to detect the moving caption patterns.

Ballantyne et al.[110] designed a new Branching Particle Filter.The filters simulate a large number of independent particles, each of which moves with the stochastic law of the target. Particles are weighted, redistributed and branched depending on the method of filtering based on their accordance with the current observation from the sequence. Each filter provides an approximated probability distribution of the target state given all back observations. The filters converge to the exact conditional distribution as the number of particles goes to infinity.

Kouritzin et al. [111] designed a nonlinear filter that addressed the distribution of present state and this model is used for invisible and random dynamic signal. This signal gives description of a distorted and corrupted partial observation operation. Kouritzin [108] also brought in a novel noise deduction algorithm called the Chopthin Algorithm. This algorithm abides by these resampling techniques until finalized output of the new particles is obtained. The discussed algorithm decreases the resampling noise.

Priyadarshi et al. [112] discussed about digital video communication systems in which a video is constructed for the purpose or aim of storing data with bit rate constrictions. The video processing is accomplished by the sum of absolute differences with the assistance of image processing block sets. The branching type particle filtering is generally used for background extraction. The authors have employed object tracking or tracing for existing time video. It has been observed that available present method can trace objects in many circumstances. The main consideration in the organization of the moving or actuating items can be employed to decrease its many-sided quality. By the image method or scheme, the capability of a computer to recover or restore the position and to recreate the object from a sequence of images is adverted. Hence, this framework employs a movement model.

Abdelali et al. [113] presented a method for object tracking based on the deterministic search of target whose color content matches a reference histogram model. A simple RGB histogram-based color model is used to develop the observation system. A new approach is described for moving object tracking with Particle Filter by shape information. In this approach Particle Filter and the probability product kernels are combined as a similarity measure using integral image to compute the histograms of all possible target regions of object tracking in video sequence. The shape similarity between a target and estimated regions in the video sequence is measured by their normalized histogram. Target of object tracking is created instantly by selecting an object from the video sequence by a rectangle.

Bauermeister et al. [114] implemented and evaluated the Sequential Importance Resampling Particle Filter. This work performs radar target track filtering with the help of the quantitative simulation. Target track filtering also called target track smoothing aims to minimize the error between a target's predicted and actual position. This objective has been achieved through the development of a software radar target filter simulator which implemented a Sequential Importance Resampling (SIR) Particle Filter algorithm and suitable target and noise models.

Zhou et al. [115] solved the problem of sample impoverishment after resampling. It is the main reason that leads to the divergence of particles and failure of profiling. To solve this problem, a new resampling algorithm is presented. The idea is that random particle samples are drawn from the neighborhoods of previous samples with high weights according to Gaussian Distribution instead of simple duplication. Therefore, during the resampling, the effect of sample impoverishment is reduced and diversity of particle samples is enriched because of sample expansion while the low weight samples are discarded.

Chou et al. [116] evaluated the performance of particle filter with the help of point estimator comparison. The Kullback-Leibler Divergence (KLD) method is basically used for this kind of particle cloud comparison. KLD estimates are generally used in particle filtering applications and it is normally applicable to whatever cloud of particles that are present. Particle Filters are commonly used for such nonlinear filtering problem. However, a major issue is to estimate the KLD for known densities only and described by a limited number of samples which might not share the same support like in particle filtering applications. The work presented a practical and generally applicable tool to do so. Particle Filter performance analysis is generally done using the mean square error between point estimators. It works fine for Gaussian cases but problems such as high order moments cannot be well approximated by a Gaussian density. In that case, another measure has to be used and the comparison of entire posterior distributions is needed.

Kumar et al. [117] considered video as 3D or 4D spatiotemporal intensity pattern i.e. a spatial intensity pattern that changes with time. In this work, analysis of the various aspects like compression and enhancement of video has been performed. There are many issues and challenges that still exist related to video processing inclusive of security issues which the author has briefly discussed in this work.

Kokaram et al. [118] presents a new and simpler Bayesian framework that achieves joint processing of noise, missing data and occlusion. It explains that in distinctive video sequences, the scene content is always principally the same from frame to frame. This theory explains that for noise reduction as well as for missing data interpolation, a lot of data needs to be synthesized to fill the gap and to bring out the underlying 'original clean' data by comparing it with the present data. Using motion compensated processing, the described pixels are extracted along motion trajectories. These motion trajectories must be approximated using one of the listed motion estimators. If the present data relates to the same statistical process, then the mean is a good estimate of the clean data. The described technique has been used to perfectly create better effect videos from electron microscope imagery.

Li et al. [119] explained advancement in sensory as well as mobile computing technology and lot of concerning applications that affect moving or actuating objects. The objects may be vehicles, ship or airplanes. No matter what the object might be or what might be the absolute volume of data and the complexity within, review of the moving objects demands a lot of human power. In this framework, object trajectories are represented using motifs fragments. Technologies like Radar and GPS devices permit trajectory data or information to be recorded for objects of all sizes such as a tiny cell phone.

2.6 MOTIVATION OF RESEARCH

Most of the existing techniques are either supervised or semi-supervised i.e., they require a data set that has been labeled as ”normal” and ”abnormal” which is then used for training a classifier. Considering the vast applications of surveillance cameras and the enormous amount of data collected by them, it seems almost impossible to label the entire data set as normal or abnormal. There is a suppressing need to develop a practical anomaly detection system which is unsupervised and runs in real time so as to timely signal an activity that deviates from the normal patterns.

Most existing methods are patch or trajectory-based which lack semantic understanding of scenes and may split targets into pieces. They are also time consuming. To handle this problem there is need for a novel and effective algorithm that incorporates deep object detection and video profiling with full utilization of spatial and temporal information.

Most of the existing intelligent anomaly detection algorithms exhibit poor performance because of the inconsistent appearance of pedestrians owing to posture deformation and clutter in video streams. Furthermore, partial occlusions significantly increase the difficulty of anomaly detection in case of a single camera view. Partial occlusions and posture deformation restrict the ability of multi-object tracking while capturing the feature representations of moving objects, especially in cluttered environments, which has not been explored yet.

Motivated by these research gaps, an online anomaly detection method for surveillance videos based on video profiling using the Bayes Filters has been proposed.

2.7 OBJECTIVES

The following are the objectives of proposed research work:

i. To develop a new probabilistic behavioral model based on a Hidden Markov Model.

ii. To develop a novel visual behavior modeling approach which performs incremental and adaptive behavior model learning for online abnormality detection.

iii. To analyze the color information and shape or appearance extraction & motion cue to achieve the image segmentation.

iv. To develop a model for the dimension reduction of the system to reduce the cost and complexity of the system for the improvement of efficiency.

The above objectives are implemented in MATLAB R2016a.
The following objectives have been achieved in the current study:

i. The proposed technique develops a visual behavior modeling approach by introducing the concept of video profiling for online abnormality detection. Use of Bayes Filters have been made to perform profiling of moving objects in videos. These filters are based on the Hidden Markov State Space Model which predicts the posterior probability based on prior belief and observation measurements. The observable variables (observation process) are related to the hidden variables (state-process) by some functional form.

ii. For abnormality detection, the process of profiling requires the extraction of foreground image from the background. This helps in detecting the moving object. The proposed technique achieves this by making use of the frame differencing and the feature extraction method.

iii. Once the moving object is identified, it needs to be located in each frame. To achieve this objective use of various feature descriptors like the HOG, Optical Flow, color features etc. has been made. These techniques provide useful color, texture and motion information and edge extraction features to achieve image segmentation and profiling.

iv. The last objective has been achieved by using dimension reduction technique i.e. The Harris Corner Detector. This technique detects corner points in the image plane. Instead of performing operations on each pixel, processing needs to be done at only these corner points thereby reducing the dimensions and making the system computationally less complex and time consuming.

2.8 CHAPTER WISE ORGANISATION OF THESIS

An outline of this thesis is presented as follows:

Chapter 1: Introduction

This chapter provides a brief introduction to video processing: its techniques and applications, activity recognition and anomaly detection in videos and various approaches for detecting anomalous pattern in videos.

Chapter 2: Literature Review

This chapter provides in detail the relevant literature survey which has been carried out to understand the various aspects relating to background estimation, foreground detection, feature extraction, and activity recognition. Different anomaly detection techniques analyzed by various researchers have been included in this chapter. The chapter also discusses the proposed objectives of the current study. In the end the chapter wise organization of the thesis has been discussed.

Chapter 3: Modified Kalman Filter Using Global and Local Oriented Gabor Texture Histogram

This chapter includes the tracking of objects in video using the Kalman filter tracking. Interest points are detected using Harris Detector. Thereafter the GLOGTH feature descriptor has been deployed to identify the target object in each frame before applying Kalman filter for tracking. Programming in MATLAB has been used to accomplish the objective.

Chapter 4: Modified Particle Filter Using Global and Local Oriented Gabor Texture Histogram

This chapter deals with the detection of abnormal behavior using particle filters. In this analysis, feature points have been extracted in the image. The object of interest has been identified and located in each frame of the video sequence by applying the GLOGTH and color feature descriptor, at these points. Tracking using particle filters has been performed to predict the position of the object in each frame. Any deviation between the observed and predicted value has been considered an anomaly. Programming in MATLAB has been used to accomplish the objective.

Chapter 5: Branching Particle Filter Based Anomaly Detection

This chapter deals with finding abnormal or unusual activities in videos using the concept of branching particle filters. The performance of the proposed techniques has been analyzed on the basis of various performance parameters used in tracking. The results obtained thereafter have been compared with the earlier techniques. Programming in MATLAB has been used to accomplish the objective.

Chapter 6. Conclusion and Future Scope.

This chapter deals with the overall conclusions of the current study and explores the future scope of work related to the field.

2.9 SUMMARY

To counter-act increasing and more diversified threats in today's societies, governments and other organisations are turning their focus towards mass surveillance. In certain circumstances, it is necessary to analyse the behaviour of people and vehicles in order to determine what is happening in a scene. More specifically, it can often be vital to determine if an event is normal or an anomaly.

Traditional passive surveillance is proving ineffective as the number of available cameras for an operator often exceeds the operators ability to monitor them. Furthermore, monitoring surveillance cameras requires a focus that operators can only uphold for a short amount of time.

To improve security and the performance of surveillance systems, it is thereby necessary to construct an algorithm that can help operators to perform at a higher efficiency or even take over the work altogether. Such algorithms would then allow the system to operate and interact in real-time with its surrounding, instead of just as a forensic tool. The algorithm might also increase the chances of detection as it can use the moment pattern of individuals or crowds that can otherwise be hard to distinguish by the human eye. However, in order for the algorithm to be reliable in a real world scenario, it is necessary that it has a low amount of false anomaly detections as well as a low amount of undetected anomalies.

This chapter expresses the work done in the past in the field of object segmentation, detection and tracking of objects in video surveillance. An extensive research work has been done for detection of video objects. Object detection in dynamic backgrounds is now an essential part in video surveillance applications. In addition, video object segmentation is a primary task for different computer vision applications. However, video object segmentation based on multi features such as size, color, texture, shape, intensity etc has not been efficiently performed in the past. The key aim of this research is to locate the moving objects exactly and then to track them for detection of any anomalous behavior. By tracking the object across the frames, the normal behavior pattern is drawn and any deviation from the normal trajectory or path is then considered as an anomaly. The literature review directs the research to address the existing works limitations leading to further investigation.

Chapter 3

MODIFIED KALMAN FILTER USING GLOBAL AND LOCAL ORIENTED GABOR TEXTURE HISTOGRAM

3.1 INTRODUCTION

Visual Surveillance system [120] is used for mostly security sensitive areas. In object tracking algorithm, automated video analysis is required. The basic applications of such type of technique in high quality videos, high powered computers and video analysis. Anomalies in videos can be of different types depending upon the situation e.g. a bicycle on a pedestrian footpath or a car from opposite direction in a one-way traffic. Violent behavior inside an elevator or Automated Teller Machine (ATM) is also a kind of anomalous activity. To detect anomalous activities first the object or target needs to be detected. Object detection in videos involves verifying the presence of an object in image sequences and possibly locating it precisely for recognition. Then the detected object is to be tracked. Object tracking is to monitor an object's spatial and temporal changes during a video sequence including its presence, position, size, shape etc. These two processes are closely related because tracking usually starts with detecting objects while detecting an object repeatedly in subsequent image sequence is often necessary to help and verify tracking.

In this chapter, object tracking using Kalman Filter [121] involves two steps. In the first step, detection of moving object is performed and in the second step, frame to frame tracking of the detected object is done. Different techniques can be applied to detect moving objects in the video depending upon whether the background is static or dynamic. For static background, the frame difference technique is applied and for dynamic background, the feature extraction method has been made use of. Once the moving object in the foreground is detected, the next step is to track the object frame to frame. In this chapter, Kalman Filter combined with GLOGTH feature descriptor has been used for moving object tracking.

3.2 PROPOSED METHODOLOGY

Object tracking in real time is one of the most active research areas in computer vision [122]. The goal of object tracking is to estimate the locations and motion parameters of the target in a video sequence given its initial position in the first frame. Research in tracking plays a key role in understanding motion and structure of objects. It finds numerous applications including surveillance [123], human-computer interaction [124], traffic pattern analysis [125], recognition [126] and medical image processing [127]. Although, object tracking has been studied for several decades and numerous tracking algorithms have been proposed for different tasks, it remains to be a very challenging problem.

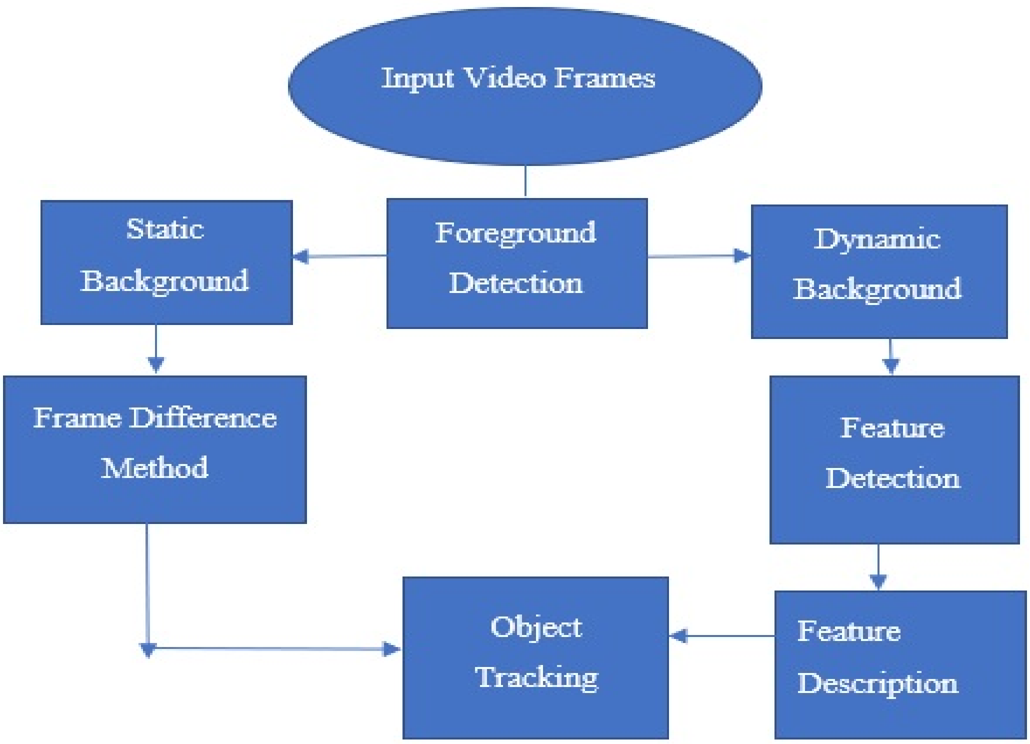


Figure 3.1: Work Flow for Moving Object Detection and Tracking

One of the most challenging factors in object tracking is to account for appearance variation of the target object caused by change of illumination, deformation and pose. In addition to these, occlusion, motion blur and camera view angle also pose significant difficulties for algorithms to track target objects. Video tracking is an application of object tracking where moving objects are located within video information. Hence, video tracking systems are able to process live, real-time footage and also recorded video files. This chapter presents a new method for real time tracking of objects using Kalman Filter. A general framework for moving object detection and tracking is given in Figure 3.1.

3.2.1 Foreground Extraction or Moving Object Detection

The first step towards achieving the desired framework is to extract the object of interest i.e. to distinguish the moving object from the background [128]. This process is called Foreground Extraction or Moving Object Detection. Moving object detection is the task of identifying the physical movement of an object in a given region or area. Over last few years, moving object detection has received much of attraction due to its wide range of applications like video surveillance, human motion analysis, robot navigation, event detection, anomaly detection, video conferencing, traffic analysis and security. In addition, moving object detection is very consequential and efficacious research topic in field of computer vision and video processing since it forms a critical step for many complex processes like video object classification and video tracking activity. Consequently, identification of actual shape of moving object from a given sequence of video frames becomes pertinent. However, task of detecting actual shape of object in motion becomes tricky due to various challenges like dynamic scene changes, illumination variations, presence of shadow, camouflage and bootstrapping problem. Number of techniques are prevalent to detect the moving object of interest in a given video sequence. Some of them are:

(a) **The Frame Difference Method** [129] is an algorithm to identify an object's motion. Using this algorithm, one can differentiate an object moving in an environment with the static background. This method adopts pixel-based difference to find the moving object. The block diagram depicting the working of frame difference method is given below in Figure 3.2.

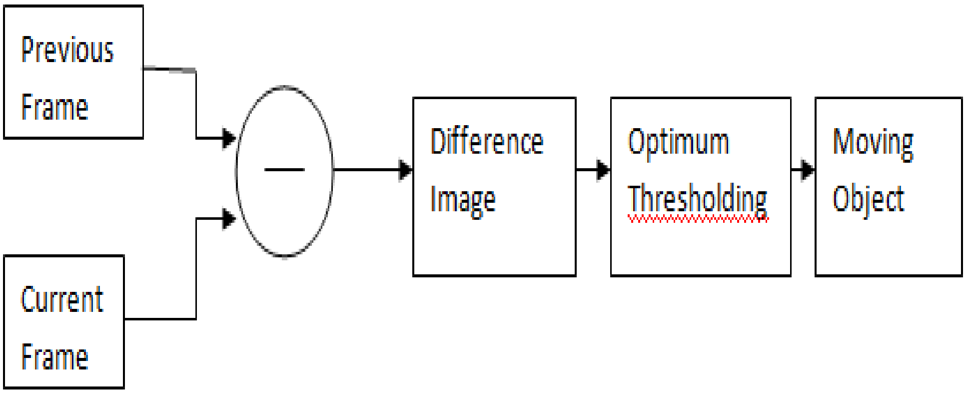


Figure 3.2: Frame Difference Method

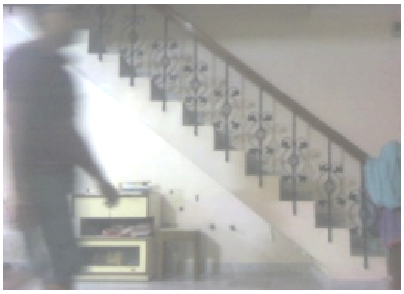
As observed from the above diagram, the current frame of the video sequence is subtracted from the previous frame. An image is obtained as a result which is the difference of the two images. Optimum thresholding is then done on the resultant image in which a threshold is decided on the basis of pixel intensities of the image. Each pixel in the image is compared with this threshold. If the pixel's intensity is higher than the threshold, the pixel is set to white in the output. If it is less than the threshold, it is set to black. Black pixels correspond to background and white pixels correspond to foreground (or vice versa). Thus, moving object in the foreground is detected [130].

Some of the problems associated with this method, especially when only using two frames are the following:

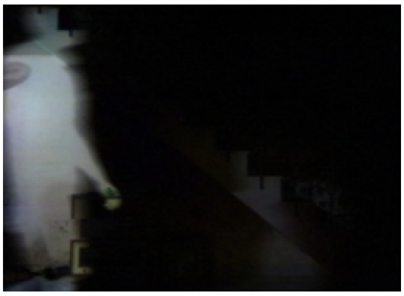
- The objects are moving slowly through the frame and only the edges of the objects are detected. This is usually referred to as the wave-front (edges) of the object being detected.
- An object that was moving becomes stationary it will disappear into the background.
- A rapidly moving object will be detected twice resulting in a ghost object at the objects previous position.

To evaluate the performance of frame difference method, take a video where a person is moving against a static background. The video is divided into frames and the frame differencing method is applied by taking the difference of the two frames. The results so obtained are shown in Figure 3.3. As is evident from the result, the frame difference method is able to detect the person moving in the foreground.

(a) Input First Frame



(b) Input Second Frame



(c) Difference Between Two Frame

(d) Binary Image of Difference

Figure 3.3: Image Showing Moving Object

This method holds the advantage of being very simple and computationally less complex but does not give reliable results when the background is dynamic, e.g. swaying of trees and movement of curtains or ripples in the river in the background. Frame difference method is sensitive to location change. The drawbacks of the frame difference include foreground aperture and ghosting caused by an object's too low motion speed. This can lead to false detection. To evaluate the performance of the frame difference technique in such a case, the method is applied on a video having a static background but having curtains swaying due to air. The video is divided into frames and the difference of the two frames is taken. The results so obtained are as shown in Figure 3.4 below:

The results clearly indicate that this technique is able to detect false movements. Though the background is static and there is no moving object in the foreground yet the results show movement because of moving curtains in the background. To remove this limitation and for foreground extraction in a video sequence with dynamic background, a technique called feature extraction is used. Feature extraction is a process of dimensionality reduction by which an initial set of raw data is reduced to more manageable groups for processing. This technique consists of two steps, first one is Feature Detection and second is Feature Description.

(a) Input First Frame

(b) Input Second Frame

(c) False Curtain Movement Detected

Figure 3.4: False Detection in Dynamic Background

(b) Feature Extraction

Feature extraction is a process of dimensionality reduction [131] by which an initial set of raw data is reduced to more manageable groups for processing. Feature extraction selects features which effectively reduce the amount of data that must be processed while still accurately and completely describing the original data set. Feature Extraction aims to reduce the number of features in a dataset by creating new features from the existing ones (and then discarding the original features). These new reduced set of features should then be able to summarize most of the information contained in the original set of features. Nearly all of the previous research was concentrated only on detection rate rather than reducing the computational complexity while maintaining high detection rate. There was need for an algorithm to ensure proper motion estimation and the detection of objects with less computation time and lower computational complexity. Feature extraction serves this purpose. It takes place in two steps:-

i. Feature Detection

ii. Feature Description.

This research uses Harris Corner Detector as the feature detector [132] and Global and Local Oriented Gabor Texture Histogram as the feature descriptor [133]. The block diagram for moving object detection and tracking using Kalman Filter is given in Figure 3.5 below.

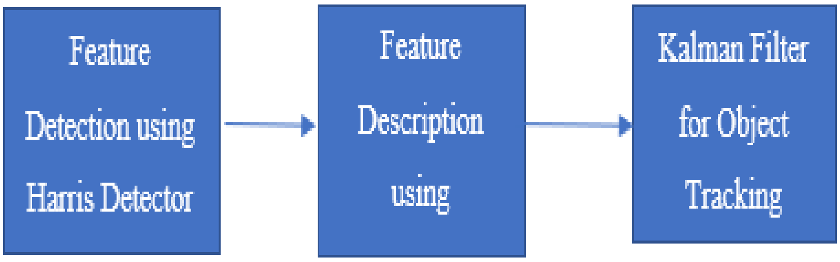


Figure 3.5: Block Diagram for Feature Extraction and Kalman Filtering

i. **Feature Detection using Harris Corner Detector** Feature is defined as an "interesting" part of an image which provides rich information on the image content e.g., corners or interest points [134]. Corners are invariant to translation, rotation and illumination. Although corners are only a small percentage of the image, they contain the most important features in restoring image information. Thus, they can be used to minimize the amount of processed data for motion tracking. Figure 3.6 shows that in a corner region, a significant gradient change is observed in all the directions.

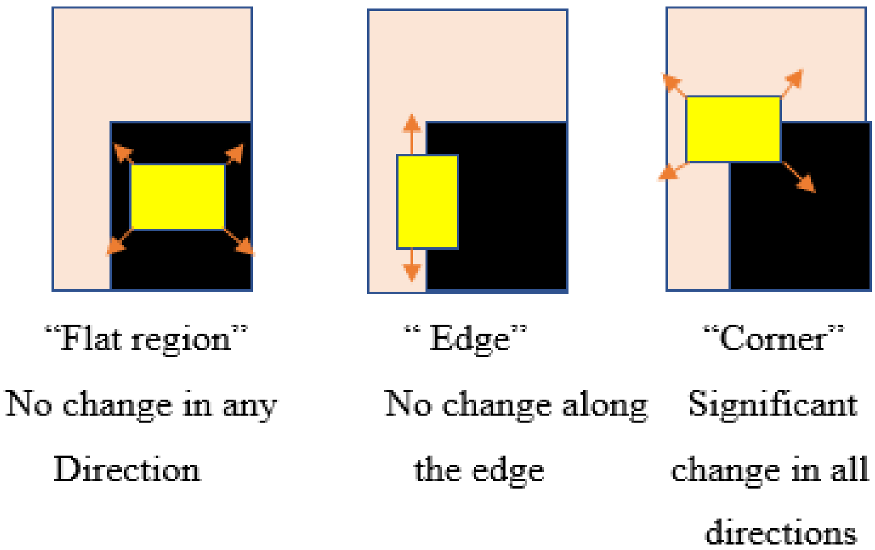


Figure 3.6: Depiction of Flat region, Edge and Corner

Harris Corner Detector

Harris Corner Detector [135] is a corner detection operator that is commonly used in computer vision algorithms to extract corners and infer features of an image. These techniques are simple and robust from computation & speed point of view. Also in terms of scaling, rotation and variation in illumination characteristics, these techniques are much better as compared to others. The Harris corner detector is therefore a standard technique for locating interest points on an image.

Despite the appearance of many feature detectors in the last decade, it continues to be a reference technique which is typically used for camera calibration, image matching, tracking and video stabilization.

The main idea is based on Moravec's Detector [136] that relies on the auto correlation function of the image for measuring the intensity differences between a patch and windows shifted in several directions. The success of the Harris detector temporary in its simplicity and efficiency. The strong variation in local neighborhood is used to locate the points using analysis of Eigen value and Eigen vector of autocorrelation matrix. This matrix, also called as structure tensor, is the base for several image processing problems such as the estimation of the optical flow [137] between two images.

From the eigen values of the autocorrelation matrix, it is possible to define several corner response functions and measures. The idea behind the Harris Method is to detect points based on the intensity variation in a local neighborhood. A small region around the feature should show a large intensity change when compared with windows shifted in any direction.

Following are the steps to calculate corner points using Harris Detector:

(a). **Color to gray scale conversion**

Convert color image into a grayscale image.

(b). **Spatial derivative calculation**

$$E(u,v) = \sum_{(x,y)\in W} w(x,y)[I(x+u,y+v) - I(x,y)]^2 \tag{3.1}$$

where $w(x,y)$ is a window function at position (x,y) that can be a rectangular or a Gaussian function.

$E(u,v)$ is the intensity variation occurred if window W is shifted by a small amount (u,v)

Intensity of shifted window is denoted by $I(x+u,y+v)$.

Intensity of original window is denoted by $I(x,y)$.

The main requirement is to maximize the value of the function E for corner detection.

$$E(u,v) \approx \sum_{(x,y)\in W} w(x,y)[I(x+u,y+v) - I(x,y)]^2 \tag{3.2}$$

Then, Taylor series expansion is done.

$$E(u,v) \approx \sum_{(x,y)\in W} w(x,y)[I(x,y) + uI_x + vI_y - I(x,y)]^2 \tag{3.3}$$

Since u and v are small, the term $I(x+u,y+v)$ approximates to $(I(x,y)+uI_x+vI_y)$ by Taylor series expansion.

$$E(u,v) \approx \sum_{(x,y)\in W} w(x,y)[u^2 {I_x}^2 + 2uvI_xI_y + v^2 {I_y}^2] \tag{3.4}$$

In matrix form the above equation may be written as:

$$E(u,v) \approx \begin{bmatrix} u & v \end{bmatrix} \left(\sum_{(x,y)\in W} w(x,y) \begin{bmatrix} {I_x}^2 & I_xI_y \\ I_xI_y & {I_y}^2 \end{bmatrix} \right) \begin{bmatrix} u \\ v \end{bmatrix} \tag{3.5}$$

Now, rename the summed-matrix as M

$$M = \sum_{(x,y)\in W} w(x,y) \begin{bmatrix} {I_x}^2 & I_xI_y \\ I_xI_y & {I_y}^2 \end{bmatrix} \tag{3.6}$$

so the equation now becomes:

$$E(u,v) \approx \begin{bmatrix} u & v \end{bmatrix} M \begin{bmatrix} u \\ v \end{bmatrix} \tag{3.7}$$

The matrix M is called structure tensor.

(c). **Corner Response Measure**

The Harris Detector uses the following response function that scores the presence of a corner within the patch.

$$R = det\, M - k(trace\, M)^2 \tag{3.8}$$

$$det\, M = \lambda_1 \lambda_2 \tag{3.9}$$

$$trace\, M = \lambda_1 + \lambda_2 \tag{3.10}$$

where λ_1 and λ_2 are the Eigen values of Matrix M. All windows that have a score R greater than a certain value are corners. They are good tracking points. Three cases are possible:

If both the eigen values λ_1 and λ_2 are small then R is small and the region is flat.

If $\lambda_1 >> \lambda_2$ or viceversa, then $R < 0$ and the region is an edge.

IF $\lambda_1 \approx \lambda_2$ and both eigen values are large, then R is large and the region is a corner.

Figure 3.7 shows how the eigen values help to determine the suitability of a window to detect corners.

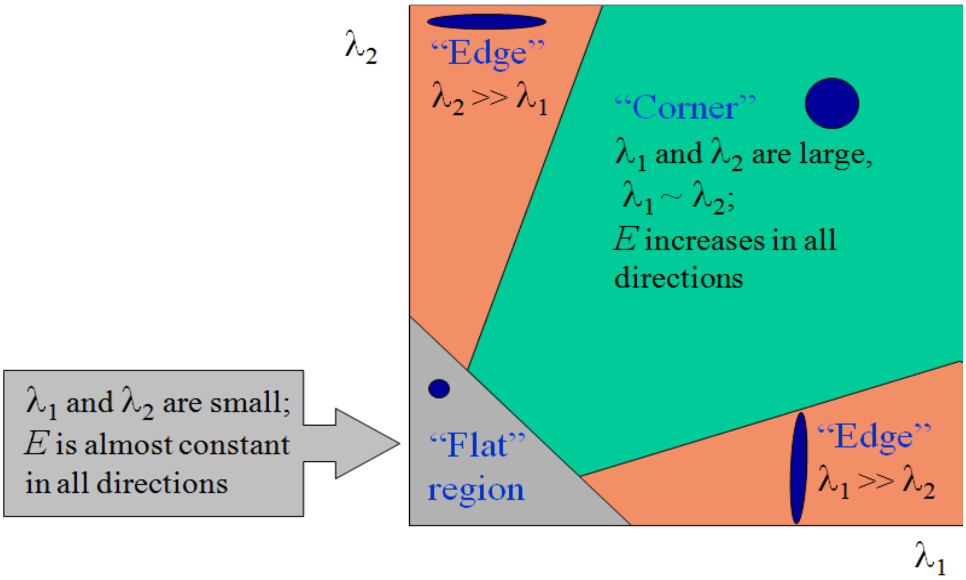


Figure 3.7: Choice of Corner Points Using Eigen Values

Based on the algorithm explained above, the Harris Corner Detector technique is applied to the image of an athlete in motion. Figure 3.8 shows the result on applying Harris Detector to the image. The feature points or corners are extracted as indicated in yellow. The significance of extracting these points in an image is that at these points there is a significant change in local information of the pixel such as its intensity, brightness, motion and appearance. Finding these cues on the entire image would lead to extracting redundant data thereby increasing computational time and complexity. On the other hand, if the cues are found out at these feature points only, lot of redundant information may be avoided thereby increasing the efficiency of the system.

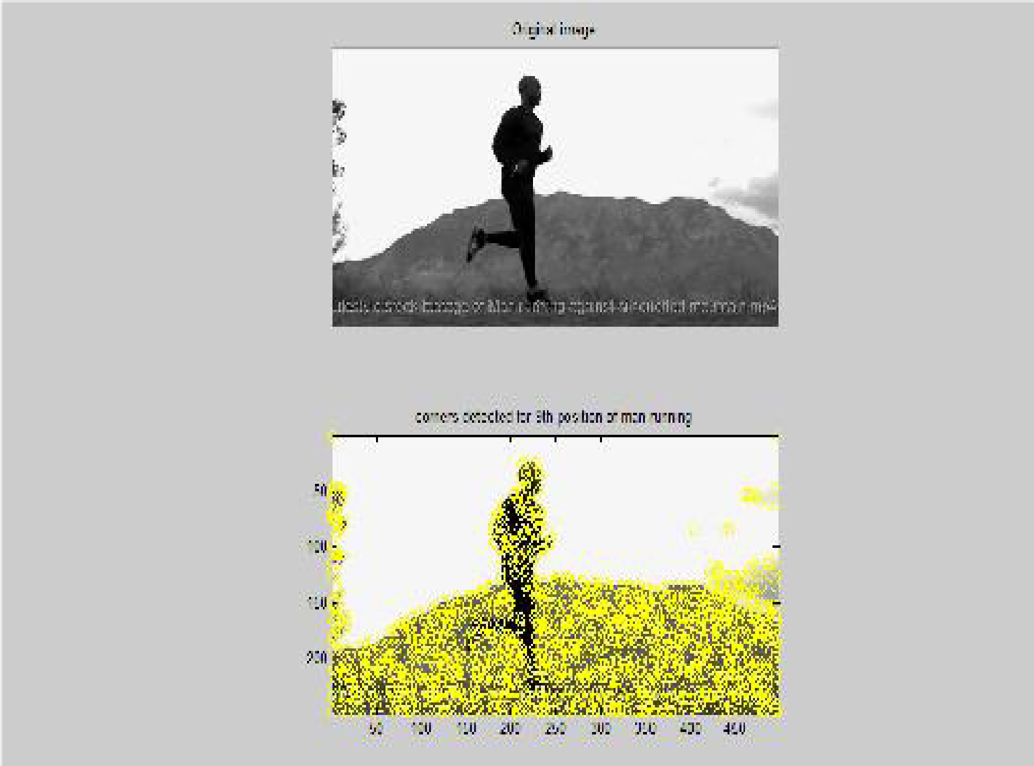
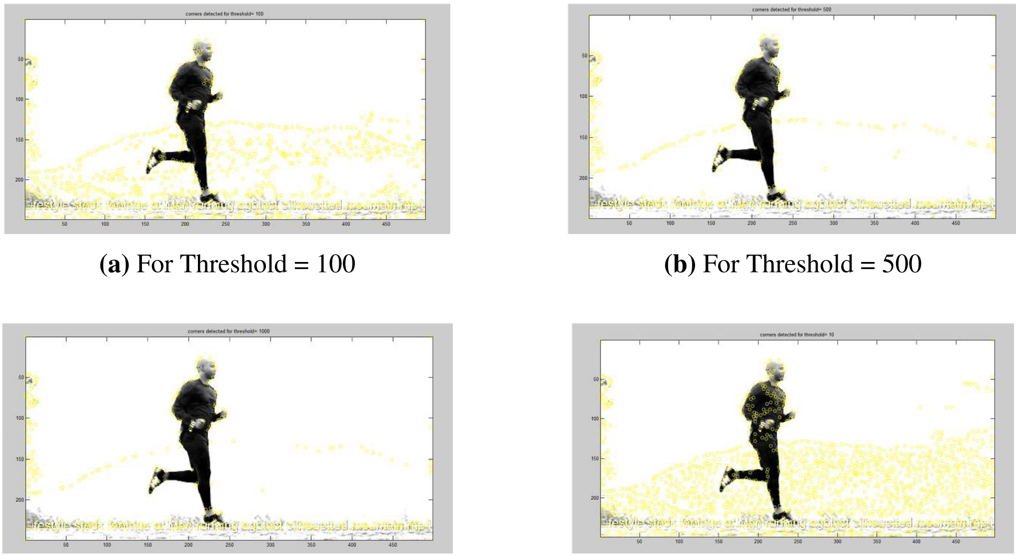


Figure 3.8: Harris Corner Detector Applied to the Image of an Athlete

Harris Corner Detector for Different Thresholds

There is no general way to define a lower threshold for the "cornerness" measure provided by the Harris Filter. It is just a score and the higher it is, the more spiky or point-like the image will look around a certain pixel. Figure 3.9 shows the importance of choosing an optimum or threshold value of the response function *R* of the Harris Detector. It can be clearly seen from the figures that when the chosen threshold value is small for e.g. 10, then a large number of corner points from the same region of image are extracted. The feature descriptor would then have to be applied on all the corner points even if they belong to the same region, thereby providing redundant information. As the threshold value increases, only a very few points are able to cross that threshold. As observed in figure 3.9 a threshold of value 1000 leads to extraction of a very few corner points which are not sufficient enough to provide the required cue information. An optimum threshold of value from 100 to 500 provides a trade-off between the two extreme cases. It is therefore of utmost importance to choose an optimum threshold value of the response function to meet both the requirements of avoiding redundant information and losing useful information.



(a) For Threshold = 100

(b) For Threshold = 500

(c) For Threshold = 1000

(d) For Threshold = 10

Figure 3.9: Harris Corner Detector for Different Thresholds

ii. **Feature Description using Global and Local Oriented Gabor Texture Histogram**

The next step towards extracting the foreground in case of videos with dynamic background is to apply a feature descriptor at those corner points which have been detected using Harris Detector. This chapter uses Global and Local Oriented Gabor Texture Histogram as the feature descriptor. Among all the local feature models, texture descriptors plays an important role in recognizing person because of its shape-free feature representation. Though variations in pose and illumination directly affect texture and shape features, the availability and potential of the texture features are excessive in the person's images. From the literature reviewed, it is noted that texture as a stand-alone feature does not work well for person re-identification and must be combined with colour or shape feature for improved performance.

GLOGTH is a combination of the local texture and global structure information of a given input image. This feature aims at representing the human appearance traits with low-dimensional feature extraction. The proposed feature extracts the texture information of input images based on the orientation of the weighted gradient from the global representation. In GLOGTH, the principal orientation is determined by the gradient of the pixels. Based on the principal orientation, the Gabor feature is extracted and imbues GLOGTH with the strong ability to express edge information, apart from making it robust to lighting variances. The essential thought behind GLOGTH [138] is that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions. It is based on the fact that Gradient (x and y derivatives) of an image is large around edges and corners (regions of abrupt intensity changes). Hence, edges and corners pack in a lot more information about object shape than flat regions. Therefore, the gradient is specifically computed at these points on the image. Using this concept, the person in the foreground can easily be detected and segregated from the rest of the scene.

GLOGTH Descriptor Algorithm:

The goal of texture analysis is to extract local region information from the image. For global feature extraction, information is extracted from the whole region of the image. In order to build rotation invariant features possessing distinctiveness, a feature is proposed by applying the concept of calculating the gradient-based structure to describe the distribution of the holistic orientations and to compensate for the global-level rotation. The Gabor-based texture describes the distribution of the local texture for each orientation. Once the global representation has been completed by representing the gradient orientation and magnitude of each pixel, the image is divided into four equal divisions of equal size i.e. upper right, upper left, lower right and lower left. The image is divided into 4 blocks, each of which is further divided into 32 cells sized 8×4. After dividing the image into cells, the gradient orientation from each pixel is extracted from the image and the orientation will be calculated for the pixels from every block. For each block, 32 orientations are calculated and ordered in the histogram. The orientations ordered to the histogram are weighed based on their gradient magnitude. From every cell the principle orientation is calculated based on the highest weight of the histogram. Other orientations are ordered based on the principle orientations. All the orientation bins of the histogram are shifted until the principal orientation places its first position.

Thereafter, the Gabor filter is applied using the 32 gradient orientations of the histogram. The Gabor feature is calculated for 4 scales and 32 orientations using the algorithm shown below.

1. The complete image is divided into small connected regions named as cell and histogram of gradient directions for each cell is computed within the

cell for each pixel. This is known as edge orientation.

2. According to gradient orientation, each cell discretizes into angular bin function.

3. The angular bin function assigned corresponding weighted gradient for each pixel cell.

4. The spatial region is assigned as grouping of adjacent cells named as block. On the basis of normalization & grouping of histograms cell is termed as block.

5. The block histogram represents normalized group histogram. A complete set of these blocks is represented by descriptor.

To evaluate the performance of the GLOGTH algorithm, the technique is applied to dataset 1. The result so obtained is shown in Figure 3.10.

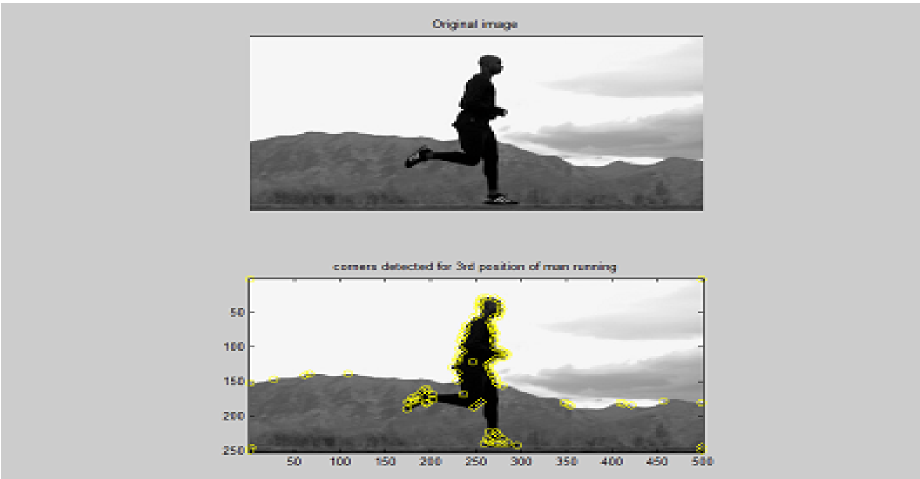


Figure 3.10: Target Identification by applying GLOGTH on Dataset 1

3.2.2 Object Tracking using Modified Kalman Filter

Once the object of interest has been detected and identified, next step is to track the object continuously over the sequence of frames. Object tracking is the problem of estimating the positions and other relevant information of moving objects in image sequences. This chapter deploys the Kalman Filter to track the position of the target object.

Rudolph E. Kalman introduced the Kalman Filter [139] in 1960. The KF, also known as the linear quadratic estimator is used for the estimation of measurement of series of data having different accuracies and noise. Kalman Filter has many uses including applications in control, navigation and computer vision. This filter takes current measurements of position, updates estimate of state variables and then predicts future. It does so iteratively i.e. it relies on the last estimate to predict the present one.

A system model used by Kalman Filter controls all the input values that would be corresponding to the updated model. A number of unknown variables are measured by measurement of series experiment with optimization techniques. These unknown variables are changed with respect to time. It is the best approach for measurement. Only single arrangements are required for this purpose. The issue occurred for discrete data linear filtering is resolved by Kalman Filter as provide a recursive solution. Kalman Filter is widely used in many applications as it is self sufficient value. In unknown system, estimation of state of system is more accurately as compared to other techniques. It is well concluded that unknown state is well estimated by Kalman Filtering. Kalman Filter is basically used for mixing accessible data measurement output which would be used by more powerful recursive estimator. The measured sensors are used as knowledgeable algorithm that would be used for estimation of unknown and the Mean Square Error (MSE) would be minimized.

Basically, such type of filter is used in linearization measurement model and known system as it is simplest linear state space model. Such types of filters are widely used in day to day life. There is uncertain value for measurement in measured value as unknown variables are estimated recursively with respect to time. In uncertainty, noise is to be added and it is included in the measured value. As estimated noise is Gaussian in nature, it cannot be measured so easily. Hence, a new algorithm or technique is required to estimate these unknown variables. This can be estimated using Kalman Filter. The linear equations and adaptive white Gaussian noise are used as standard models for Kalman Filter.

Model for Kalman Filter

For estimation & calculation of unknown states, two steps are used by Kalman Filter.

1. Prediction

2. Update

1. **Prediction**

   The Kalman filter works in two steps (prediction and correction step) that are executed one after another in the loop. In the prediction step only the known action is used in a way that enables prediction of the state in the next time step.

From the initial belief a new belief is evaluated and the uncertainty of the new belief is higher that the initial uncertainty. In this method, the current value is estimated from previous predicted value. A priori state estimation is also known as prediction state estimation as no measurement value is required. Prediction step is shown as:-

$$\hat{x}(t|t-1) = F(t)\hat{x}(t-1|t-1) + B(t)u(t) \tag{3.11}$$

$$P(t|t-1) = F(t)P(t-1|t-1)F'(t) + Q(t) \tag{3.12}$$

where,

$\hat{x}$: Estimated State.

F: State Transition Matrix (i.e., transition between states).

u : Control Variables.

B : Control Matrix (i.e., mapping control to state variables).

P : State Variance Matrix (i.e., error of estimation).

Q : Process Variance Matrix (i.e., error due to process).

Subscripts are as follows: $t|t$ current time period, $t-1|t-1$ previous time period and $t|t-1$ are intermediate step.

The measurement is done with above two equations. These are used as estimated value of priori value & error measurement respectively.

2. **Update**

In this step, accuracy of estimate step is improved by updating apriori state by observation of different values. After the estimation, observations are known as posterior estimation. Update estimation is written as:

$$e(t) = y(t) - H(t)\hat{x}(t|t-1) \tag{3.13}$$

$$R_{ee}(t) = H(t)P(t|t-1)H'(t) + R(t) \tag{3.14}$$

$$K(t) = P(t|t-1)H'(t){R_{ee}}^{-1}(t) \tag{3.15}$$

$$\hat{x}(t|t) = \hat{x}(t|t-1) + K(t)e(t) \tag{3.16}$$

$$P(t|t) = [1 - K(t)H(t)]P(t|t-1) \tag{3.17}$$

where

e : Residual measurement error

R_{ee} : Prediction covariance

y : Measurement variable

H : Measurement matrix (i.e., mapping measurements onto state).

K : Kalman gain.

R : Measurement variance matrix (i.e. error from measurements).

The most important concepts when using the Kalman filter are summarized as:

- Kalman Filters are discrete. They rely on measurement samples taken between repeated but constant periods of time.
- Kalman filters are recursive. Its prediction of the future relies on the state of the present (position, velocity, acceleration etc). Further, it presents a guess about external factors that may affect the situation.
- Kalman Filters predict the future. This is applied by making measurements (such as by sensors) and then deriving an adjusted estimate of the state from both prediction and measurements.

The Kalman Filter has several applications in technology. Some common applications are:

a. Guidance and navigation of vehicles, particularly aircraft and spacecraft.

b. Robotic motion planning and trajectory adjustment.

c. Position awareness radar sensors for advanced driver assistance systems (ADAS) in autonomous vehicles.

d. Many computer vision applications such as stabilizing depth measurements, object tracking (e.g. faces, heads, hands), fusing data from laser scanners, stereo cameras for depth and velocity measurements, and 3D mapping through Kinect or range cameras.

The next section illustrates the performance of the proposed technique on two different datasets using MATLAB as the simulation tool.

3.3 SIMULATION RESULT

This research work makes use of MATLAB 2016a as the simulation tool. To evaluate the performance of the proposed algorithm using Kalman Filter, the technique is implemented on two datasets consisting of two video sequences, dataset 1 comprising of a person jogging at uniform speed against a static background and dataset 2 comprising of cars moving at nonuniform speed agaist a dynamic background.

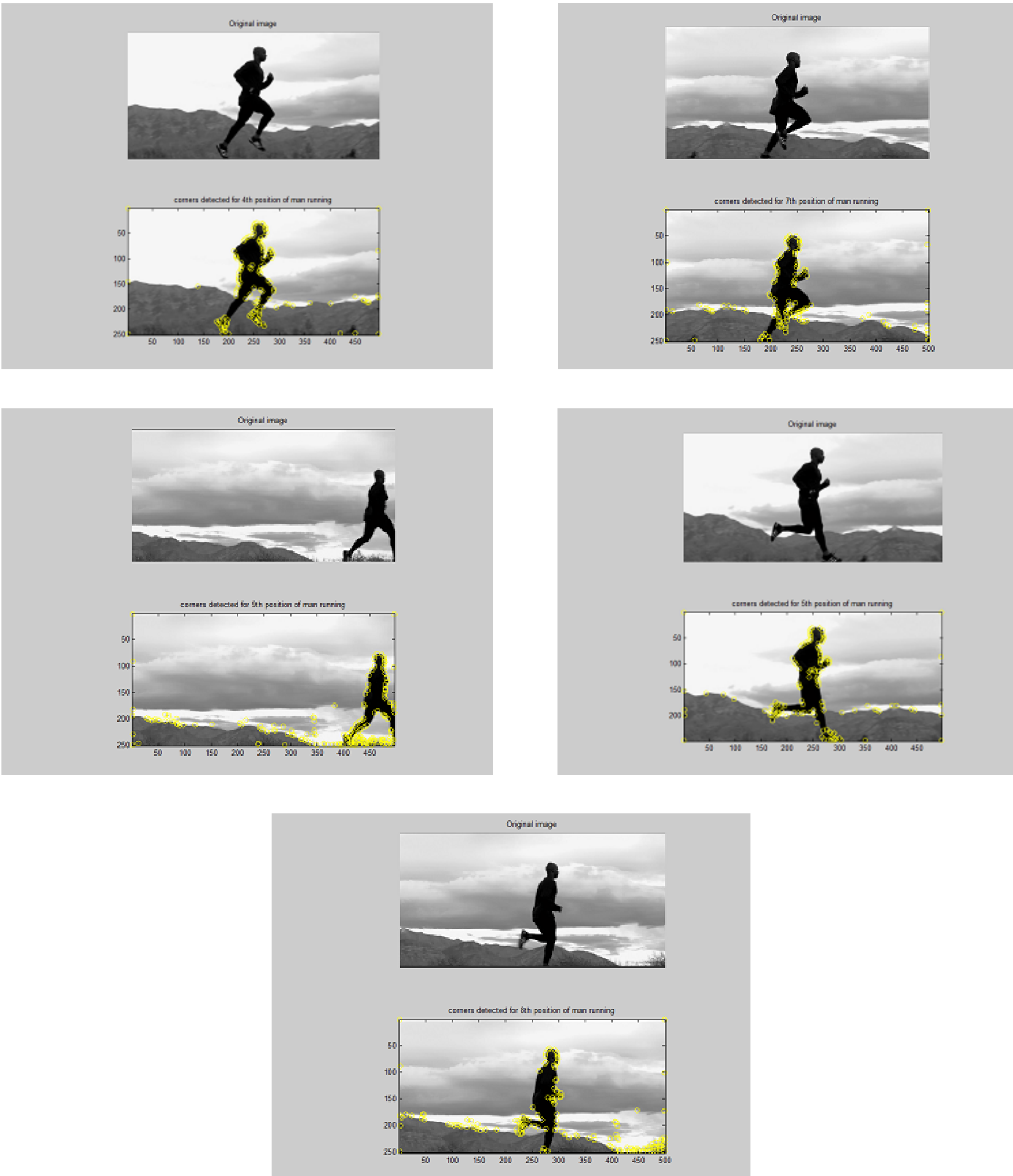


Figure 3.11: Successful Tracking of the Target on Applying Kalman Filter (Dataset 1)

Figure 3.11 shows the results obtained after applying Kalman Filter to target object for dataset 1 where the background is static and also the target object is not occluded by any other object. As is evident from the figure, the Kalman Filter is successfully able to track the target object in all the consecutive frames thereby proving that Kalman Filters are successful in tracking uniform guassian targets.

Performance Metrics

In order to evaluate the performance of our algorithm the following performance metrics have been taken into consideration:

1. **True Positives (TP)**
   It is used to measure correct predicted positive values. It means actual class is present and the predicted class is also present. It is illustrated with an example as if wolf was present and the shepherd shouted wolf.

2. **True Negatives (TN)**
   It is used to measure correct predicted negative values. It means actual class is not present and the predicted class is also not present. It is illustrated with an example as if shepherd shouted no wolf and actually no wolf was present

3. **False Positives (FP)**
   It is used to measure both actual & predicted class. It means actual class is not present and the predicted class is present. It is illustrated with an example as if wolf was not present but shouted wolf present.

4. **False Negatives (FN)**
   It is used to measure both actual & predicted class. It means actual class is not present and the predicted class is present. It is illustrated with an example if shepherd shouts no wolf when actually the wolf was present.

   The performance parameter like Recall Rate, Precision Rate, Accuracy & F Score are calculated using TP, TN, FP and FN

5. **Accuracy**
   Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. Accuracy is the fraction of predictions our model got right. The performance of filter & algorithm is measured by Accuracy. It is defined as the ratio of correctly predicted class to the total number of classes or observations.

$$\text{Accuracy} = \frac{\text{TN+TP}}{\text{TN+FN+TP+FP}} \tag{3.18}$$

6. **Precision**

   It may be defined as the ratio of correctly predicted positive class to all predicted positive classes. It means correct fraction of detected items.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{3.19}$$

7. **Recall (Sensitivity)**

   It may be defined as the ratio of correctly predicted positive class to all actual present classes. It means correct fraction of all the detected items.

$$\text{Recall Rate} = \frac{\text{TP}}{\text{TP+FN}} \tag{3.20}$$

8. **F score**

   It may be defined as average of weighted recall rate & precision rate. It means both false & true classes are measured. Since in terms of performance F score is more important than accuracy for random class distribution. If the cost function of FN & FP is same then accuracy is more dominant as compare to F score. If the cost function of FN & FP is different than recall rate & precision rate is more dominant as compare to other parameters

$$\text{F Score} = \frac{2 \times \text{Recall Rate} \times \text{Precision Rate}}{\text{Recall Rate} + \text{Precision Rate}} \tag{3.21}$$

9. **Root Mean Square Error**

   RMSE measures how much error there is between two data sets. In other words, it compares a predicted value and an observed or known value.

$$\text{RMSE} = \sqrt{\frac{\sum((\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2)}{N}} = \sqrt{\frac{\sum_{i=1}^{N} distance_i{}^2}{N}} \tag{3.22}$$

   N is the number of samples. $\hat{x}_i$ and $\hat{y}_i$ are the predicted coordinates and x_i and y_i are the known or observed coordinates.

10. **Detection Rate**

    It is a measure of the percentage of true targets that is detected. It may be defined as the ratio of number of outliers detected to the total number of outliers in data.

To evaluate the performance of the proposed algorithm, three parameters namely precision (P), recall (R) and F score (F) have been taken into consideration. With the proposed algorithm using Kalman Filter combined with GLOGTH feature descriptor, the values of Precision, Recall and F score obtained are 0.9972, 0.9601 and 0.7259 respectively. To validate the results obtained, the proposed algorithm has been compared with the work done by Komagal et al. [140] in Table 3.1. Komagal et al. uses HOG features with Hungarian Algorithm to obtain efficient tracking in wide area. Histogram of Oriented Gradients is used to obtain foreground/background. The segmented output is then applied to Hungarian Algorithm together with the 'Kalman Filter' for tracking.

Table 3.1: Performance Comparison of GLOGTH Kalman with Hungarian Kalman on Dataset1

Methods	Precision	Recall	F Score
Classical Kalman [140]	0.9964	0.9571	0.7100
Hungarian Kalman [140]	0.9969	0.9596	0.7122
GLOGTH+Kalman (proposed)	0.9972	0.9601	0.7259

To give a clear insight of the performance of the proposed technique over other algorithms, Figure 3.12 displays the data in Table 3.1 in the form of bar graphs.

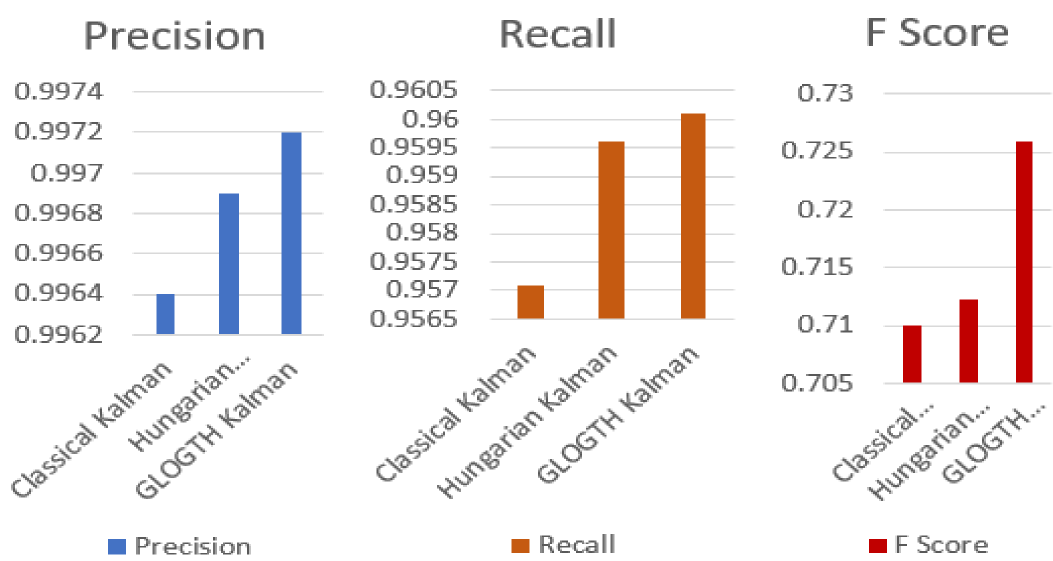


Figure 3.12: Comparative Analysis of Modified Kalman Filter

The above figure clearly indicates that the proposed algorithm outperforms both the classical Kalman and the Hungarian Kalman algorithm. This is because the classical HOG descriptor is sensitive to rotation transformation. On the contrary the GLOGTH descriptor is consistent and robust to any type of image transformations. This greatly improves the performance of the proposed work.

Now the Kalman Filter is applied on Dataset 2 in which the background is dynamic and objective is to track the red colored cars which are moving at abrupt speeds. The results so obtained are shown in Figure 3.13.

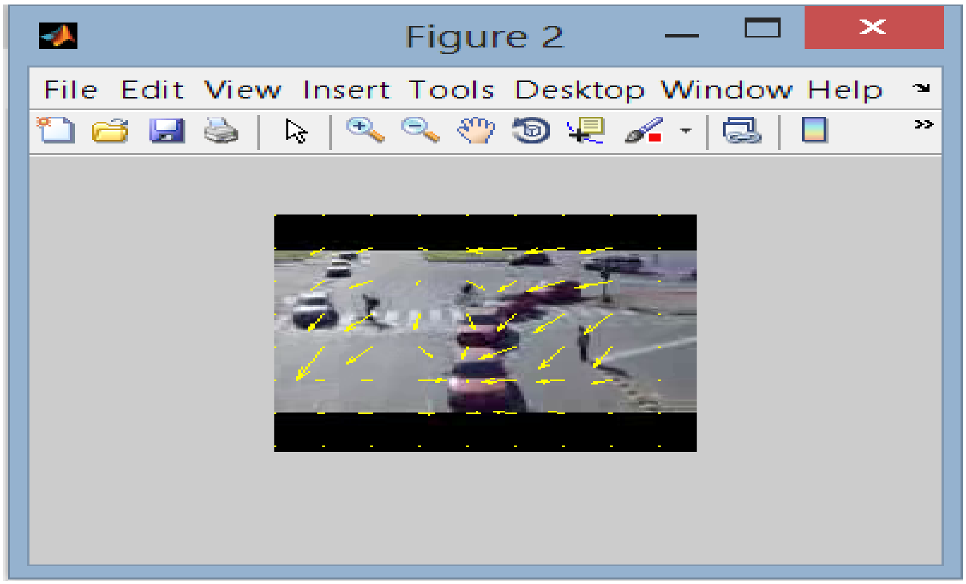


Figure 3.13: Poor Tracking Result in Nonlinear environment (Dataset 2)

The random directions of the tracker clearly indicate that the Kalman Filter fails to track the red colored cars because their speed is not uniform and direction is nonlinear.

3.4 RESULTS AND DISCUSSION

Object detecting and tracking has a wide variety of applications in computer vision such as video compression, video surveillance, vision-based control, human-computer interfaces, medical imaging, augmented reality and robotics. It also plays an important role in video database such as content based indexing and retrieval.

This chapter deploys the Kalman Filter combined with GLOGTH feature descriptor for tracking the object of interests. Before tracking, the Harris Corner Detector is used to detect the feature points in the image. These are the points where there is maximum change in gradient in all directions. By detecting these feature points, the objective of dimensionality reduction is achieved. Now instead of applying the feature descriptor at all the points in the image, only corners are taken into consideration. This drastically reduces the computation time of the algorithm. Figure 3.8 and Figure 3.9 show the application of Harris Detector to the dataset. Next, a feature descriptor is applied at all the corner points. GLOGTH is used here as a feature descriptor. GLOGTH focuses on the shape or appearance of the object in the image. Figure 3.10 depicts how when

GLOGTH is applied at the corner points, only the athlete i.e., the object of interest is highlighted. This enables the tracker algorithm to now focus and track only the target object. Figure 3.11 shows how when the Kalman Filter is applied, the tracking of the target object is achieved successfully. Next, the same techniques are applied to the second dataset consisting of cars moving at nonuniform speed in random directions. The haphazard yellow arrows of Figure 3.13 show how the Kalman Filter fails to track the cars of second dataset. This is because the Kalman Filter assumes linear Gaussian environment. Thus, when the Kalman Filter is applied to nonlinear non-Gaussian dataset, the tracker fails.

3.5 CONCLUSIONS

It is an important task to reliably detect and track moving objects for video surveillance and monitoring. Tracking can be defined as the problem of estimating the trajectory of an object in the image plane as it moves around the scene. Tracking of moving objects based on the results from object detection is aimed to estimate the optimal trace of the moving objects for further event analysis. With object tracking solutions, one can perform meaningful actions on visual data obtained via different types of cameras. Using suitable object detection algorithms coupled with tracking models, one can train a machine to not just recognize one or more unique objects or persons in a particular image but also identify them in subsequent frames and follow their trajectory in a video stream.

Moving object detection and tracking is the major challenging issue in Computer Vision, which plays a vital role in many applications like robotics, surveillance, navigation systems, militaries, environmental monitoring etc. There are several existing techniques which has been used to detect and track the moving object in Surveillance system. Tracking objects in a video sequence can be complex. The difficulty can arise due to rapid appearance change caused by image noise, illumination, non rigid body motions or because of dynamic backgrounds, occlusions and interaction between multiple objects. Loss of depth information caused by mapping objects and points from 3D space to a 2D image plane is also a challenge. Therefore, it is necessary to develop new algorithm or modified algorithm which is robust and sturdy.

In this chapter the conventional Kalman Filter is combined with the GLOGTH feature descriptor and applied on two different datasets. The results obtained after applying Kalman Filter to the datasets lead to following conclusions:

i. Kalman filter works extremely well for linear models such as those where there are no abrupt changes in speed or direction of the tracked object.

ii. Kalman Filter leads to poor tracking performance in environments which introduce nonlinear factors such as occlusion or multipath propagation effects [141]. Moreover, if there is only one single moving object, the tracking problem is trivial. However, if there are multiple moving objects sometimes overlapped (or occluded) with each other, the tracking problem become more difficult.

Chapter 4

MODIFIED PARTICLE FILTER USING GLOBAL AND LOCAL ORIENTED GABOR TEXTURE HISTOGRAM

4.1 INTRODUCTION

Most of the algorithms for object detection and tracking [141] are able to track the particular object only in controlled and predefined conditions. In video processing & computer tracking, when the target is moving at nonuniform speed or is changing its direction abruptly then it becomes a challenging task to track and find out the exact location of the targets.

Moreover, most algorithms do not take into account the removal of redundant data. In this chapter, a robust model for object tracking has been developed which uses the Harris Detector Technique for feature point extraction. Now RGB color model has been merged with the GLOGTH feature descriptor [142] to obtain better results and Particle Filters [143] have been deployed to enhance the accuracy and performance parameter of tracking. Particle Filters provide robust tracking by working in nonlinear non-Gaussian environment [144] and feature extraction process helps in reduction of data dimensionality.

Experimental results of the proposed tracking system show that this system is robust in nature as it models the uncertainty in motion of the object under consideration and shows improvement in the recognition and tracking performance compared to other techniques like Kalman Filter [145], Mean Kalman Filter [146], Unscented Kalman Filter [147].

Any object tracking system involves the following sequence:

i. Feature Point Extraction or Feature Detection

ii. Feature Description

iii. Object Tracking

Here, Harris Detector has been used for corner point detection, RGB color model plus GLOGTH for feature description and Particle Filter for object tracking.

4.2 PROPOSED TECHNIQUE

As discussed in Chapter 3, the biggest limitation of Kalman Filter is that it cannot be used in nonlinear non-Gaussian environments and also fail to detect and track objects which are occluded by other objects. To overcome this limitation, use of Particle Filters has been made for object tracking. Before this, the process of feature extraction is applied to detect moving objects in the foreground. This feature extraction process takes place in two steps namely, Feature Detection and Feature Description. The Harris Detector is used to extract feature points in the image. The RGB color model is then combined with GLOGTH [148] is used to identify the target object in subsequent frames based on appearance and color. The application of the Harris Detector reduces data dimensions thereby reducing the computation time of the algorithm. The generalized framework of the proposed system is shown in Figure 4.1.

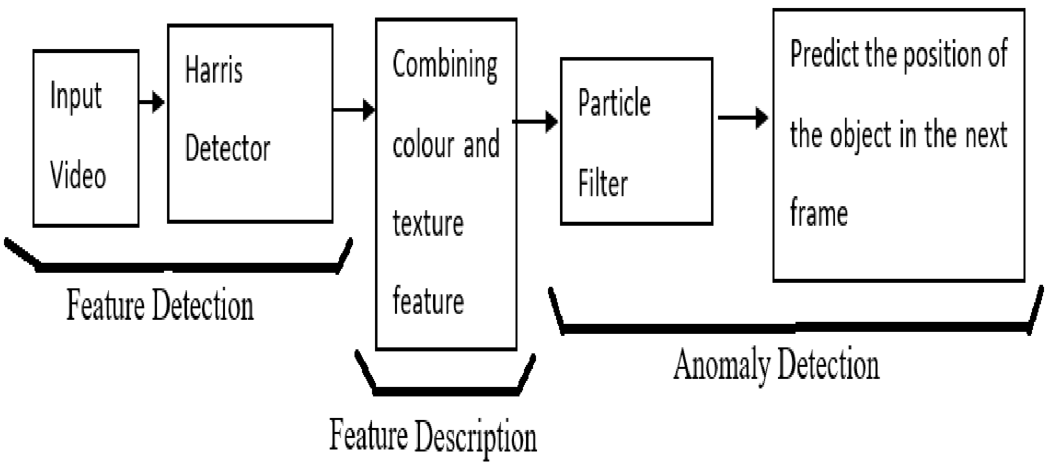


Figure 4.1: Anomaly Detection using Particle Filter

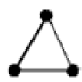
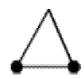
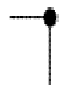
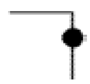
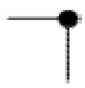
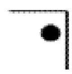
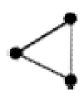
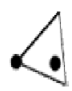
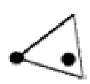
4.2.1 Feature Point Extraction or Feature Detection

Traditionally, the term detector has been used as a tool that extracts the various image features from the images. A wide variety of feature detectors exist e.g. corner detector, blob detector or edge detector [149]. These detectors are used to identify the local points or features. The detected attributes or features are useful in image processing & computer detection e.g. objects recognition, matching, alignment of images and mosaicking applications.

Corners are the most local features in the images. In the video frames, corner points play a major role as the variation in pixel intensity is maximum at these points. Moreover, these points are invariant to rotation angle change, scale change, change of visualization angle and illumination changes. The operation of corner point extraction can be used for data processing minimization without any loss of data.

The most common feature detectors are SUSAN [150], Kitchen-Rosenfeld [151], Harris, KLT [152] and FAST [153]. In above mentioned intensity-based techniques, Harris Corner Technique [135] is the best suitable algorithm for feature detection.

Table 4.1: Requirements of a Corner Detector

Sr. No.	**Requirements**	**Good Corner Detectors**	**Bad Corner Detectors**
1.	Detection of all ”True Corners”		
2.	No Detection of ”False Corners”		
3.	Localization of all Corner points		
4.	High Repeatability Rate i.e., Better Stability of Detector		
5.	Robust nature of detector w.r.t noise		

Corner detectors have to satisfy several criteria as shown in Table 4.1. First, all true corners should be detected. Second, no false corners should be detected. Third, the corner points should be well localized. Fourth, the most important property of a corner detector should be its high repeatability rate. Fifth, the corner detector should be robust with respect to noise and should be computationally efficient.

The Harris Detector

Features & interest points are calculated by a mathematical operator technique named as Harris Corner Detector Algorithm. [135]. For reducing the computational complexity and computation time and increasing the speed of the algorithm, this technique is simple and robust. Due to the invariance of the detector to scaling, rotation and variation in illumination changes, this technique is much better as compared to others. Thus, the Harris corner detector outperforms the standard techniques for locating interest points on an image.

4.2.2 Feature Description

The visualization features of image and video are expressed by feature descriptors. Elementary characteristics like motion, texture and color of the object of interest in an image or video sequence are defined by the feature descriptors. It can be defined as a technique that takes an image as input and outputs its features in the form of special feature vectors.

Several cues can be used for feature description which provides a series of information that can help to distinguish the object of interest from other objects present in the scene. The information obtained from the feature description is invariant even if images are transformed. The most elementary features covered by the common descriptors used in image and video processing are color, texture, shape, motion and location. Some most commonly used feature descriptors are SIFT [154], SURF [155], MSER [156], ORB [157], HOG, and HOF [158].

The RGB color model combined with GLOGTH has been used as the feature descriptor.

i. **RGB Color Descriptor**

   Color is a basic feature for image representation and is invariant to the scaling, translation and rotation of an image. The human eye is sensitive to colors and color features are one of the most important elements enabling humans to recognize images. Color features are fundamental characteristics of the content of images. Color features can sometimes provide powerful information for categorizing 64 images and they are very useful for image retrieval [159]. Therefore, color-based image retrieval is widely used.

Color is what different objects reflect and their wavelengths entering one's eyes depict the type of color one can see. Color is the characteristic of ethnic visual perception described through color categories including names such as red, yellow, purple and blue. This understanding of coloration derives from the excitement concerning cone cells among the human sight using electromagnetic radiation in the spectrum of light. Color categories and physical specifications of colour are also associated with objects, materials, light sources etc. based on their physical properties such as light absorption, reflection and emission spectra. The colors perceived of objects are the results of interactions between the various frequencies of visible light waves and the atoms of the materials that objects are made of. Many objects contain atoms capable of either selectively absorbing, reflecting or transmitting one or more frequencies of light. The frequencies of light that become transmitted or reflected to one's eyes will contribute to the color that one perceives.

Color detection in image processing [160] is detecting a particular colored object in the image. In the case of videos, color detection is applied to the frames taken from the video. Tracking a particular colored car in traffic can be of great benefit and will detect the object separating the rest of the background from that particular color or neglecting the rest. But color detection has got nothing to do with motion. It will detect the particular colored object whether in motion or stationery.

For applying color detection to a video frame, first of all, the video frames need to be converted into a grey scale image from the colored image. Then the grey sale image is subtracted from the colored image depicting dimensions of that particular color that needs to be detected such as to extract red component from RGB image, one can use the function 'rgbimage(:,:,1)'. The resultant image holds only that colored component which is to be detected. But, due to subtraction some noise gets introduced into the image which is to be removed by using a noise filter such as a Median Filter. Now the grey scale image is converted to the binary image. This has a benefit as luminance points greater than are assigned value 1 and luminance points less than one are assigned value 0. To detect all red colored objects in an image, the labels of all the connected objects in the image need to be collected under one heading for further operation.The number of components that can be connected under one label can be 4, if mentioned or by default, the value assigned is 8. This is done by using the MATLAB function 'bwlabel'. The next step is to match the dimensions of the labels of the connected objects with connected properties i.e. position, dimensions etc. Once the position of that particular colored object is obtained, mark the object on the image (colored one) by drawing a shape according to the size of the object whichever suits best. The shape can be a circle or rectangle. This is done by using the MATLAB function 'drawcircle' or 'drawrectangle'. The program is dynamic as it can detect any color on that image by simply changing the dimensions of the color that needs to be detected and can be modified to detect any color of choice.

ii. **Global and Local Oriented Gabor Texture Histogram Feature Descriptor**
Global and Local Oriented Gabor Texture Histogram or GLOGTH is a feature descriptor that is often used to extract features from image data. It is widely used in computer vision tasks for object detection. The GLOGTH descriptor focuses on the structure or the shape of an object. In the case of edge features, one can only identify whether the pixel is an edge or not. GLOGTH can provide the edge direction as well. This is done by extracting the gradient and orientation (or one can say magnitude and direction) of the edges.

Additionally, these orientations are calculated in 'localized' portions. This means that the complete image is broken down into smaller regions and for each region, the gradients and orientation are calculated.

Finally, the GLOGTH would generate a Histogram for each of these regions separately. The histograms are created using the gradients and orientations of the pixel values. Therefore, the algorithm has been named 'Global and Local Oriented Gabor Texture Histogram'.

The basic steps of the process for implementation of the GLOGTH descriptor are-

a. **Pre-processing the data**
Preprocess the image and bring down the width-to-height ratio to 1:2. The image size should preferably be 64×128. This is because the image will be divided into 8×8 and 16×16 patches to extract the features.

b. **Calculating Gradients**
The next step is to calculate the gradient for every pixel in the image. For this take a small patch from the image, get the pixel values for this patch

and generate the pixel matrix for the given patch as shown in Figure 4.2.

121	10	78	96	125
48	152	68	125	111
145	78	85	89	65
154	214	56	200	66
214	87	45	102	45

Figure 4.2: Pixel Matrix Generation

To determine the gradient (or change) in the x-direction, subtract the value on the left from the pixel value on the right. Similarly, to calculate the gradient in the y-direction, subtract the pixel value below from the pixel value above the selected pixel. The same process is repeated for all the pixels in the image.

c. **Calculate the Magnitude and Orientation**

Determine the magnitude and direction for each pixel value. Apply the Pythagoras Theorem to calculate the total gradient magnitude

$$\text{Total Gradient Magnitude} = \sqrt{\left[(\mathrm{G_x})^2 + (\mathrm{G_y})^2\right]} \tag{4.1}$$

where G_x and G_y are gradients in x and y directions respectively. Figure 4.3 shows the total gradient magnitude.

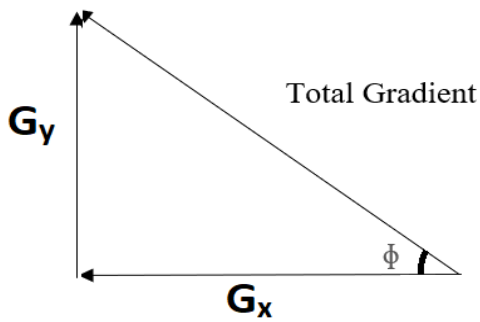


Figure 4.3: Total Gradient Magnitude

Next, calculate the orientation (or direction) for the same pixel using the formula

$$\tan(\Phi) = \frac{G_y}{G_x} \tag{4.2}$$

Hence, for every pixel value, the total gradient (magnitude) and the orientation (direction) have been obtained. Next, it is needed to generate the histogram using these gradients and orientations.

d. **Create Histograms**

A histogram is a plot that shows the frequency distribution of a set of continuous data. Take the angle or orientation on the x-axis and the frequency on the y-axis. To generate histograms, take each pixel value, find the orientation of the pixel and update the frequency table. The process has been explained in Figure 4.4 by taking an example for the highlighted pixel (85). Since the orientation for this pixel is 36, add a number against the angle value 36, denoting the frequency of that angle value.

121	10	78	96	125
48	152	68	125	111
145	78	85	89	65
154	214	56	200	66
214	87	45	102	45

Frequency						1									
Angle	1	2	3	4...	35	36	37	38	39...	175	176	177	178	179	180

Figure 4.4: Histogram Generation

The same process is repeated for all the pixel values and it ends up with a frequency table that denotes angles and the occurrence of these angles in the image. This frequency table can be used to generate a histogram with angle values on the x-axis and the frequency on the y-axis.

e. **Normalize Gradients**

Since different images may have different contrast, contrast normalization is done on the histogram vector within a block using the formula in Equation (4.3).

$$\text{L2-norm}: f = \frac{v}{\sqrt{|v|_2^2 + s^2}}; \quad \text{L1-norm}: f = \frac{v}{|v|_1 + s}; \quad \text{L1-sqrt} = \sqrt{\frac{v}{|v|_1 + s}} \tag{4.3}$$

where, v be the non-normalized vector containing all histograms in a given block, $|v|_k$ be its k-norm for $k = 1,2$ and s be some small constant.

Finally, a descriptor is assigned to each detector window. This descriptor consists of all the cell histograms for each block in the detector window. The detector window descriptor is used as information for object recognition [161].

4.2.3 Object Tracking

In computer vision tools, object tracking is one of the important domains of research analysis. The exact location of moving objects at varying times is done by information technology using different rolling cameras. The tracking of moving objects and videos can be analyzed using many frames. There are some applications of object tracking like editing of videos [162], traffic controlling [163], image/video compression, tracking, surveillance, augmented reality [164] etc. A lot of data is contained in the video which makes video tracking a time taking process as compared to image processing. The target objects are to be tracked in the consecutive video frames. This is the main objective of video tracking. When objects are moving fast as compared to frame rate then it is difficult to achieve tracking. The second challenge is when orientation changes with respect to time. In such complex cases, one can define a motion model which would describe the change in motions of objects for video tracking systems. In video tracking, different techniques are used for analyzing frames of output target and proposed video frames. A lot of algorithms are used to track the video. Each algorithm has its advantages and disadvantages. In a visual tracking system, following techniques [165] are used as a major component of the process:

i. **Target definition and localization**

   To identify the moving objects, many methods are used as a tool for targeting the objects. The computational time & complexity for such a technique is the least. These techniques are further clssified into following types of algorithms:

   a. Kernel-Based Tracking (Mean-shift Tracking) [166]

   The mean-shift algorithm is an efficient approach to tracking objects whose appearance is defined by histograms (not limited to only colors)

   b. Contour Tracking [167]

   It is also known as border following or boundary following. Contour tracking is a technique that is applied to digital images in order to extract their boundary. This method iteratively progress a primary contour in the previous frame to its new position in the current frame.

ii. **Filtering and Data Association**

These techniques are used for evaluating different hypotheses, starting information of objects or scenes and the moving speed of the object. These algorithms are used in complex motion object location & detection as they track the objects behind obstructions.Thus their computational time & complexity are much higher than other algorithms.The following algorithms are used for filtering:

a. Kalman Filter

These filters are more effective for detecting motion. A linear function named as Bayesian function is used as a filter for removing Gaussian noise. It is a much more efficient algorithm, as in a single measurement, one can estimate the number of unknown variables that are produced from noise and inaccuracies of measurement. It can be operated for a long duration of time.

b. Particle Filter

This filter is used for non-Gaussian noise and nonlinear systems for state space distribution systems.

Particle Filter for Object Tracking

There is various problem arising in image processing. Due to this Bayesian statistical interference occur in the image and signal. To resolve these problems, Sequential Monte Carlo (SMC)methods or Particle Filters [168] are used. In dynamical systems, partial observations are made due to estimating the internal states creating filtering problems and the presence of random perturbations in dynamic systems and sensors. The main objective of the applied Markov Process is to calculate noise and some observations for removing noise. In the same way, particle filtering is used to collect different sample for the distribution of processes that produces noise. The noise distribution, non-linearity and partial observation of particles can be defined by the state space model.

For state distribution or state space model, there are some assumptions required for generating the sample. This can be done using the Particle Filtering Technique. The predication correction methods are used for updating in an approximation manner using Particle Filters. In this method, a set of particles are used as sample distribution. Each particle has the same weight defined by the probability density function. In such types of filtering techniques, weight collapse occurs due to weight disparity. It can be resolved using resampling. In the resampling technique, new particles are introduced on negligible weight particles on the other hand higher weights are replaced by those in proximity.

The optimal solution to the approximate model is provided by a Particle Filter rather than that of the exact model provides an approximate solution. The derivation of the state space model is derived using the Monte Carlo Technique i.e. part of Particle Filtering. In each & every step, noise estimation is measured in the state of the system. For such a dynamic system, the state-space model can be described as follows:

$$x_t = a(x_{t-1}, u_t) \tag{4.4}$$

and the measurement model is given by:

$$y_t = b(x_t, v_t) \tag{4.5}$$

where hidden states x_t and measurement data y_t are assumed to be generated by nonlinear functions $a()$ and $b()$ respectively of the state noise process u_t and measurement noise v_t.

Steps of Particle Filtering

There are four steps for particle filtering initialization, time update, measurement update and resampling:

(a). **Initialization**

In this step consider that N random particles are assigned and all assigned particles have equal weights. Consider that object is moving in the two-dimension plane and a particle moving in the plane is having both the co-ordinate points. The parameter of the model is changed according to the position and speed of the particle.

(b). **Time Updat**e

For achieving target dynamics, the particle sets are propagated in a particular direction according to the process model. This process occurs after particle set initialization.

For each Particle, the first step is a prediction

$$x_{t+1}^{i} = f_t(x_t^i) + w_t(x_t^i) \tag{4.6}$$

where x_{t+1}^{i} is the Bayesian estimate at time $t+1$, f is a non linear time-varying function for the state x_{t+1}^{i}, $w_t(x_t^i)$ is a mutually independent sequence with a known probability density function and is called the importance weight. As the growth of particles increases, approximation error decreases.

(c). **Measurement Update**

For updating the particle weight value, the Probability Density Function (PDF) is used which is based on measurement techniques. The weights are adjusted using the measurement.

Weights before adjustment

$$\hat{w}_{t+1}^{i} = w_t^i p\left(\frac{y_t}{x_t^i}\right) \tag{4.7}$$

$p\left(\frac{y_t}{x_t^i}\right)$ is the filtering probability density.

Weights after adjustments

$$w_{t+1}^{i} = \frac{\hat{w}_{t+1}^{i}}{\alpha} \tag{4.8}$$

$$\text{Here } \alpha = \sum_{i=1}^{N} \hat{w}_{t+1}^{i} \tag{4.9}$$

In the above equations, it is observed that the particles will lose weight when they are far away from the true object state. Similarly, the particles will gain weight when they are near the true object state.

(d). **Resampling**

Compute an estimate of effective number of particles as

$$N_{eff} = \frac{1}{\sum_{i=1}^{N} \left(w_{t+1}^{i}\right)^2} \tag{4.10}$$

when N_{eff} falls below a certain threshold value the resampling process is done. Since the resampling process is performed in each case in a particle filter, accuracy increases. Normally $\frac{2N}{3}$ or $\frac{3N}{4}$ threshold values are taken into account to process.

Several approaches are used for resampling. Out of these approaches, systematic resampling is most commonly used due to certain advantages. In this technique, particle weights are updating their speed and position every time accordingly. After updating this information, a random number is generated for every new sample. The process is repeated again & again so that value is updated regularly. The process is repeated 100 times to find the exact path of the object. In the first iteration, particle weights are going to $\frac{1}{N}$ times of original weights.

The random number is generated when the resampling technique is applied in such cases. In such cases, only a single valued function is obtained for getting an optimal solution as a single particle is picked up. After some iterations, the particles are again subdivided into subsets. The particles go to an estimated state if further divided into more parts.

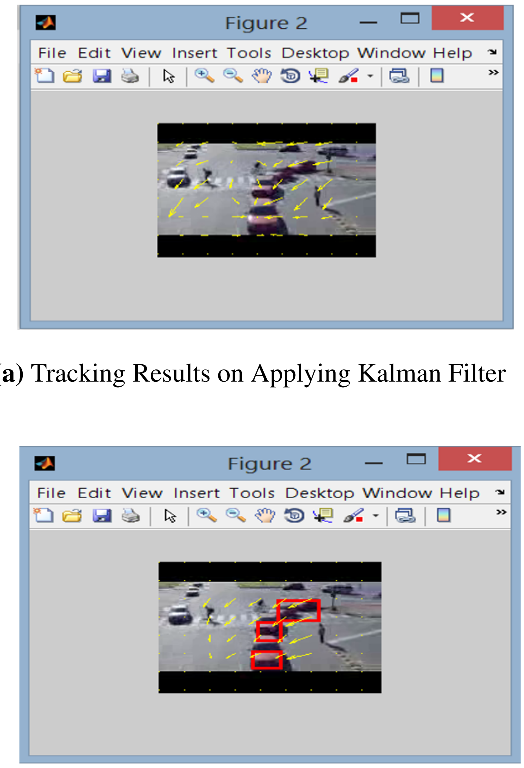
This condition is worst when particles go into an estimated state. To avoid this state, the mean is calculated so that it will go to the new estimated state of an object. In the current scenario, extensive research has been carried out based on this concept.

4.3 SIMULATION RESULTS

In this section, the performance of the proposed modified particle filter is evaluated by using MATLAB R2016a as a working platform and applying the technique on the Dataset 2. This dataset comprises of a video in which multiple cars are moving at non uniform speed against a dynamic background and the aim is to detect and track all the red colored cars.

Figure 4.5a shows the failure of Kalman Filter in tracking multiple objects moving at non uniform speed. Figure 4.5b shows the successful tracking of the red colored cars on appying modified Particle Filter. The figures indicate that in the dataset where the Kalman filter failed to detect cars moving at nonlinear speed and was changing direction abruptly, the Particle Filter was successful in tracking the red-colored colors.

The performance evaluation of the proposed technique is carried out by calculating the values of precision, recall and F score and then by comparing these values with the work carried out by Komagal et al. [140]. Table 4.2 compares the values of precision, recall and F score of the proposed technique with that of the classical Kalman Filter and the Hungarian Kalman Filter [140] as shown below.



(a) Tracking Results on Applying Kalman Filter

(b) Tracking Results on Applying Particle Filter

Figure 4.5: Tracking Results of Kalman and Particle Filter

Table below compares the proposed technique which combines Particle Filter, RGB color and GLOGTH feature descriptor with the Kalman Filters. The proposed work shows more accuracy as it is not only capable of handling partial occlusions and poor illumination environments but is also robust when the texture features of vehicles are similar. The algorithm proposed by Komagal et al.[140] shows poor performance under these conditions.

Table 4.2: Performance Comparison of GLOGTH Particle Filter with Hungarian Kalman on Dataset 2

Methods	**Precision**	**Recall**	**F Score**
Classical Kalman [140]	0.9682	0.9714	0.4772
Hungarian Kalman [140]	0.9689	0.9696	0.4872
GLOGTH+Particle Filter (proposed)	0.9711	0.9724	0.5121

Figure 4.6 gives the comparative analysis of the modified Particle Filter in the form of a bar graph. The above figure indicates that the proposed work combining the Particle Filter with the GLOGTH feature descriptor outperforms the classical Kalman and the Hungarian Kalman Filter proposed by Komagal et al.[140].

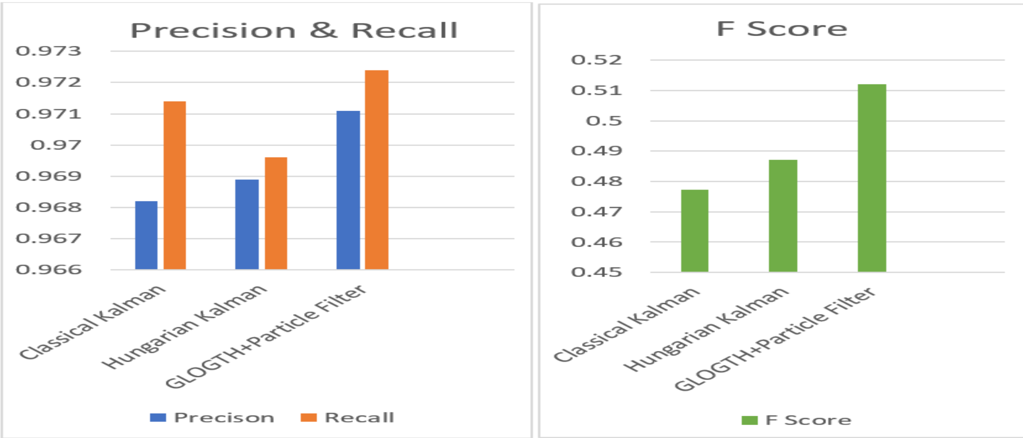


Figure 4.6: Comparative Analysis of Modified Particle Filter

Komagal et al.[140] modified the Kalman Filter by combining it with the Hungarian algorithm. This removed the drawback of classical Kalman Filter of not being able to track objects in the nonlinear non-Gaussian environment yet it suffers from the drawback of using additional tracking points in the moving object region which serve as a representative of each object or vehicle that is detected. The use of these additional points makes the algorithm take more computational time and also reduces its efficiency and accuracy.

The proposed Particle Filter on the other hand combines the virtues of three versatile techniques namely, the Particle Filter, the GLOGTH Feature Descriptor and the Color Feature Descriptor. Particle Filters show good results in nonlinear non- Gaussian environment. They model the partially available information from samples. This enables the tracker to track the objects even in the case of partial occlusion. Color features are robust to changes in orientation and size of image and GLOGTH is robust to illumination and pose variation. The three techniques when combined give excellent result in tracking multiple objects moving at nonuniform speed against dynamic background where there might be a probability of partial occlusion.

The Modified Particle Filter though overcame the limitations of Kalman Filter by successfully tracking moving objects in a nonlinear non-Gaussian environment, still they suffer from some major drawbacks. The Particle filters fail to track objects that are fully occluded by other objects. This possibility of the object of interest getting occluded by another object is highest in video datasets with high dimensionality that is in video datasets that capture crowded scenes. Moreover, they require a higher number of particles to represent the posterior volume of the object state. Thus, they suffer from heavy computing cost.

(a) Original Video Dataset Depicting Crowded Scenario

(b) Tracking Results Obtained on Applying Particle Filter to Track People Wearing Red Colored Clothes

Figure 4.7: Original video and tracking results of modified particle filter

Also, at each resampling step, noise particle are also resampled. This leads to degradation of original signal. In Particle Filters, the samples with high weights are carried forward while those with negligible weights are neglected. After a few iterations the entire weight is concentrated on only a few particles. Since, the prediction success of Particle Filters depends on the number of particles, therefore, few leftover particles decreases the success rate of the Particle Filter. This problem is called the sample impoverishment or the particle degeneracy problem [169].

The results obtained by applying Particle Filter to a crowded video dataset are shown in Figure 4.7a and 4.7b.

The above dataset captures a crowded scenario where the aim is to track people wearing red-colored clothes. The results obtained on applying the Particle Filter to the above dataset indicate that Particle Filters are only able to track two people wearing red-colored clothes as they are not occluded by any other person or object. It fails to track the third person who has been partially hidden because of the black-colored bag and has also been occluded by other people. Thus, it can be said that particle filters fail to provide successful tracking in crowded scenarios.

4.4 RESULTS AND DISCUSSION

This chapter deploys the Particle filter for tracking the target object. Before applying the tracker algorithm, the Harris detector is again deployed to achieve dimensionality reduction. Now instead of utilizing only the GLOGTH feature descriptor, this chapter deploys GLOGTH combined with a color descriptor to extract features and identify the target object. The red squares in Figure 4.5b depict how GLOGTH combined with color descriptor can identify only cars and that too only the red colored cars from a scene consisting of a wide variety of objects like pedestrians, traffic poles and multicolored cars. Once the target object has been identified, the Particle Filter is deployed to track the target. The well-defined direction of the yellow arrows in Figure 4.5b shows that the particle filter can successfully track objects in a nonlinear non Gaussian environment.

Once the Particle Filters showed good tracking results on Dataset 2 the work was extended to evaluate the performance of the Particle Filter on a dataset that depicted high dimensional data such as a dense crowd of people at a fair or a historical monument where the object of interest might be partially or fully occluded by another object. Such a situation is depicted in dataset 3 as shown in Figure 4.7a. Here the aim was to track people wearing red colored clothes. The red bounding boxes in Figure 4.7b show how the color descriptor combined with the GLOGTH descriptor was able to differentiate the people wearing red-colored clothes from all the other objects in the scene but on the application of the Particle Filter to track those persons, the Particle Filter was able to track only those persons who were not occluded or were very partially occluded by any other object. This is because with the Particle Filter approach the system can model the partially available information from samples which gives a higher chance for the tracker to track the objects even in the case of partial occlusion. However, the Particle Filter failed to track the person who was fully or very much occluded by another object [170].

4.5 CONCLUSIONS

The conventional Particle Filters have been combined with GLOGTH and color feature descriptor for object detection and tracking. Particle Filter is one of the representatives of generative tracking algorithm. Particle Filters have been used widely in the tracking problem. Particle Filter algorithm has the advantage of simplicity and flexibility. It is easy to handle non-Gaussian and multimodality system model. The classical Particle Filter usually adopts the dynamic model with global information. Regardless of whether the target is blocked or deformed, it treats the target as a whole. This leads to the neglect of the local information of the target. When the target is partially occluded and local appearance of it changes, Particle Filter algorithm cannot accurately track the target. During the above analysis, it has been observed that the Particle Filters show good results in a nonlinear non-Gaussian environment thereby overcoming the limitation of Kalman Filters. Moreover the Particle Filter is combined with color feature descriptor which being scale and rotational invariant have showed robustness to partial occlusion and is computationally efficient.

The Particle Filters also suffer from certain drawbacks although they are much efficient than the Kalman Filters. If the target object is completely or much occluded by another object, then the Particle Filter fails to track the occluded target. Thus, Basic Particle Filters give poor results in tracking objects in crowded video sequences where there is a possibility of the target object getting occluded. Moreover, Particle Filter (PF) tends to suffer from heavy computing costs because it requires a higher number of particles to represent the posterior volume of the object state. They also suffer from Bootstrap Particle Problem [171]. This problem arises because noise particles are also resampled at each resampling step of the algorithm. This leads to poor performance of the filter.

Particle Filter methods involve a mandatory resampling procedure that eliminates the particles that become redundant and multiplies the ones that contribute most to the resulting approximation. After some time the complete weight is concentrated on only a few particles. This will affect our estimation or tracking process as the particles will converge too soon. This is known as sample degeneracy or impoverishment [169].

Hence, the experimental results demonstrate that the proposed algorithm can track the moving objects well under illumination changes, partial occlusion and moving background but it fails in high dimension dataset consisting of large crowds where the object of interest is fully occluded.

Chapter 5

BRANCHING PARTICLE FILTER BASED ANOMALY DETECTION

5.1 INTRODUCTION

The main objective of moving object detection techniques is to find out exact path of the object in a plane. The region of interest or object of interest in a video is tracked by moving object tracking algorithm for analyzing its behavior in an optimum manner.

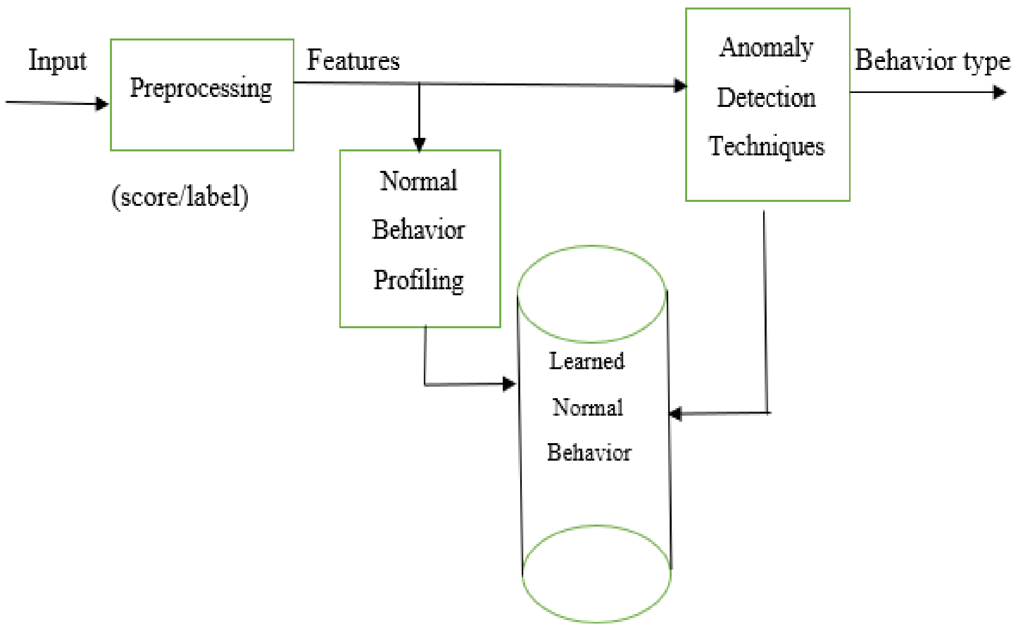


Figure 5.1: Anomaly Detection in Videos

In the tracking of moving object, there are three steps to follow. In the first step, region of interest is calculated for objects that are in motion in the video. This process is known as motion detection [1]. The second step is known as object tracking [172].

It is used to track the selected object. The analysis of movement of object to study its behavior is done with third step. The third step finds application in abnormality detection or crowd behavior detection or violent behavior detection [173]. Figure 5.1 shows the block diagram for anomaly detection in videos.

Keeping in view the vast applications of moving object tracking and detection in all walks of life, be it domestic or commercial, sports or national security, tremendous research work has been done in this field over the years. A number of methods and approaches are available to track moving objects in a video. The most commonly used being Kalman Filter, Particle Filter, MHT, Template Matching [174], Mean Shift, Cam-Shift SVM [175] and many more.

There are various merits & demerits of each technique. For stable & linear system having Gaussian noise, Kalman Filters are preferred over others. Particle Filters show good results in nonlinear non Gaussian environment thus overcoming the limitations of Kalman Filters but Particle Filters show poor results in crowded scenes with dynamic background.

As cocluded in Chapter 4, the Particle Filters suffer from a few limitations. Due to the resampling step added in Particle Filters, two problems arose. First, at each step, noise particles are also resampled leading to an increase in the noise effect on the system. Second, in resampling step, the particles with higher weight are carried forward and those with lesser weights are removed. After a few iterations it is observed that particles with higher weight simply crumple into one, leading to a decrease in the diversity of particles thereby diminishing the accuracy of approximation. This is called the problem of particle degeneracy and impoverishment [169].

The current study describes a new method of tracking a single moving object in a video using Branching Particle Filter [109, 110, 111] for providing precision in tracking. The various performance parameters like accuracy & computational time are optimized by Branching Particle Filter.

5.2 METHODOLOGY

This section elaborates the technique of detecting abnormal behavior in video frames by the process of tracking. Here the tracking has been performed by using a new class of filters called the Branching Particle Filters. These filters come under the n- class of Bayes Filter [176] which use the concept of ”prediction”. Thus, the normal activity of the object of interest is predicted in the next frame and any deviation between the observed and predicted behavior is considered as abnormal. The process of tracking involves the usual steps of moving object detection and tracking followed by anomaly detection.

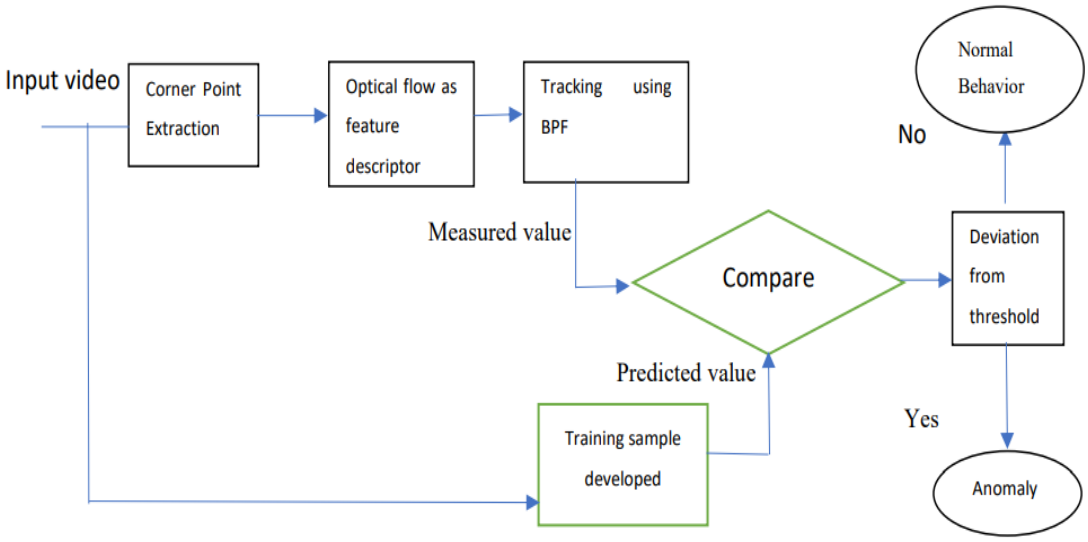


Figure 5.2: Tracking and Anomaly Detection using BPF

Steps involved in anomaly detection using object tracking

i. Moving Object Detection

ii. Feature Detection and Extraction

iii. Tracking

iv. Anomaly Detection

Figure 5.2 shows the process of tracking and anomaly detection using BPF.

5.2.1 Feature Detection using Harris Detector

The Harris Corner Detection Technique [135] has been employed to detect feature points in the images. The Harris Corner Detector is a mathematical way of determining points in the images where there is maximum variation in intensity, gradient and other local features. Feature descriptors when applied at these points show extreme variations in intensity texture and brightness etc. which can be used to identify and locate the object of interest in each frame of the video sequence. Thus, corner points are good tracking points. The detailed algorithm of Harris Detector has already been explained.[135]

5.2.2 Feature Description using Optical Flow

Optical flow is the pattern of apparent motion of various objects, surfaces and edges in a visual scene which is caused by the relative motion between an observer (an eye or a camera) and the scene . The concept of optical flow was introduced by the American psychologist James J. Gibson to describe the visual stimulus which is provided to animals moving through the world [1]. Gibson stressed on the importance of optic flow because of its ability to discern possibilities for action within the environment. Gibson's ecological approach to psychology demonstrated the role of the optical flow stimulus for the perception of movement of the observer in the world, perception of different shapes, distance and movement of objects in the world and the control of motion. The term optical flow is also used by robocist, encompassing related techniques from image processing and control of navigation which includes motion detection, object segmentation, time-to-contact information, focus of expansion calculations, luminance, motion compensated encoding and various stereo disparity measurement. Figure 5.3 depicts the concept of optical flow pictorially.

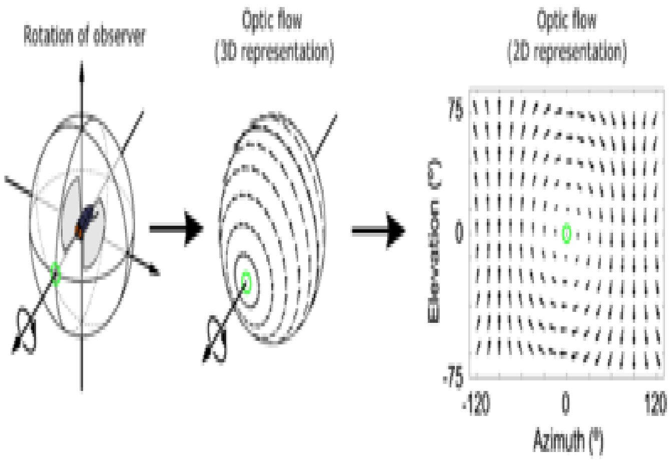


Figure 5.3: Concept of Optical Flow [1]

Optical Flow Estimation

Sequence of ordered images allows the estimation of motion as either instantaneous image velocities or as discrete image displacements. Fleet and Weiss [177] introduced the concept of gradient based optical flow. Barron, Fleet and Beauchemin [178] provided a performance analysis on a number of optical flow techniques. It emphasizes on the accuracy and density of different measurements.

For every location in image, optical flow method calculates motion between any two consecutive frames with respect to time i.e. time taken between t and t+Δt. Method uses differentials based on Taylor series expansion and uses partial derivatives in time and space coordinates of image. The intensity I(x, y, t) moved by distance be Δx, Δy and Δt between two image frames for a point in 2- dimension is given as:

$$I(x,y,t) = I(x+\Delta x, y+\Delta y, t+\Delta t) \tag{5.1}$$

Taking movement to be very small and expanding the R.H.S of Equation (5.1) using Taylor series:

$$I(x+\Delta x, y+\Delta y, t+\Delta t) = I(x,y,t) + \frac{\delta I}{\delta x}\Delta x + \frac{\delta I}{\delta y}\Delta y + \frac{\delta I}{\delta t}\Delta t + H.O.T. \tag{5.2}$$

where,
$\frac{\delta I}{\delta x}$, $\frac{\delta I}{\delta y}$ and $\frac{\delta I}{\delta t}$ are the derivatives of the image. Neglecting the higher order terms and substituting Equation (5.2) in Equation (5.1), it follows:

$$\frac{\delta I}{\delta x}\Delta x + \frac{\delta I}{\delta y}\Delta y + \frac{\delta I}{\delta t}\Delta t = 0 \tag{5.3}$$

Dividing throughout by Δt

$$\frac{\delta I}{\delta x}\frac{\Delta x}{\Delta t} + \frac{\delta I}{\delta y}\frac{\Delta y}{\Delta t} + \frac{\delta I}{\delta t}\frac{\Delta t}{\Delta t} = 0 \tag{5.4}$$

Above equation results in

$$\frac{\delta I}{\delta x}V_x + \frac{\delta I}{\delta y}V_y + \frac{\delta I}{\delta t} = 0 \tag{5.5}$$

where, V_x, V_y are the x and y components of the velocity respectively or optical flow of $I(x,y,t)$.
$\frac{\delta I}{\delta x}$, $\frac{\delta I}{\delta y}$ and $\frac{\delta I}{\delta t}$ can also be written in the form of derivatives as I_x , I_y and I_t.

Thus:

$$I_xV_x + I_yV_y = -I_t \tag{5.6}$$

This is an equation in two unknowns and cannot be solved as such. This is commonly known aperture problem in the optical flow algorithms. To find the optical flow, one more set of equations is needed which is to be given by some additional constraint. All optical flow methods introduce some additional conditions for estimating the actual flow under optical flow.

Motion Detection by Lucas–Kanade Method

In computer vision, the Lucas–Kanade method [179] is a popularly used differential method for optical flow estimation which was developed by B.D. Lucas and T. Kanade. This method assumes that the flow is essentially constant in a local neighbor-hood of the pixel under consideration and solves the basic optical flow equations for all the pixels in that neighborhood, by the least square criterion.

By combining information from several nearby pixels, Lucas–Kanade method often resolves the inherent ambiguity in the optical flow equation. It is also less sensitive to image noise than point-wise detection methods. But on the other side, since this is a purely local method, it cannot provide any flow information about objects in the interior of uniform regions of the image.

Under Lucas-Kanade method it is assumed that the displacement among nearby pixels of a image is very little and is considered as almost constant with a neighboring point. Optical flow equation should also work well for all the pixels which all are within the region of center point. The velocity vectors (V_x, V_y) must satisfy the equation-

$$\begin{aligned} I_x(q_1)V_x + I_y(q_1)V_y &= -I_t(q_1) \\ I_x(q_2)V_x + I_y(q_2)V_y &= -I_t(q_2) \\ \cdots \quad \cdots \quad & \cdots \\ \cdots \quad \cdots \quad & \cdots \\ \cdots \quad \cdots \quad & \cdots \\ I_x(q_n)V_x + I_y(q_n)V_y &= -I_t(q_n) \end{aligned} \tag{5.7}$$

where q_1, q_2, . . . , q_n are depicting pixels which are located inside the window. The equations can also be transformed into matrix form i.e., $AV = b$, in which

$$\begin{aligned} A &= \begin{bmatrix} I_x(q_1) & I_y(q_1) \\ I_x(q_2) & I_y(q_2) \\ .. & .. \\ .. & .. \\ .. & .. \\ I_x(q_n) & I_y(q_n) \end{bmatrix}, \\ V &= \begin{bmatrix} V_x \\ V_y \end{bmatrix}, \\ b &= \begin{bmatrix} -I_t(q_1) \\ -I_t(q_2) \\ .. \\ .. \\ .. \\ -I_t(q_n) \end{bmatrix}, \end{aligned} \tag{5.8}$$

As compared to the number of unknowns, the system has more number of equations. This problem is solved by taking least square criterion and it solves the matrix by multiplying the transpose of matrix A on both the sides of the equation $AV = b$.

Thus:

$$A^T AV = A^T b \tag{5.9}$$

or

$$V = (A^T A)^{-1} A^T b \tag{5.10}$$

in which A^T is the transpose matrix. It computes:

$$\begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} \sum I_x(q_i)^2 & \sum I_x(q_i)I_y(q_i) \\ \sum I_y(q_i)I_x(q_i) & \sum I_y(q_i)^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum I_x(q_i)I_t(q_i) \\ \sum I_y(q_i)I_t(q_i) \end{bmatrix} \tag{5.11}$$

where $I_x(q_i)$, $I_y(q_i)$ and $I_t(q_i)$ are the partial derivatives of the image I with respect to position x, y and time t evaluated at the point q_i and at the current time and the summation is running from i=1 to n.

Weighted Window

Under least square criterion all pixels are given equal importance or weightage but an edge or more weights should be assigned to pixels which are near to the central pixel. To deal with that, a weighted version of least square criterion is created.

$$A^T WAV = A^T Wb \tag{5.12}$$

or

$$V = (A^T WA)^{-1} A^T Wb \tag{5.13}$$

In which W is weight diagonal matrix of $n \times n$, which computes

$$\begin{bmatrix} V_x \\ V_y \end{bmatrix} = \begin{bmatrix} \sum w_i I_x(q_i)^2 & \sum w_i I_x(q_i)I_y(q_i) \\ \sum w_i I_y(q_i)I_x(q_i) & \sum w_i I_y(q_i)^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum w_i I_x(q_i)I_t(q_i) \\ -\sum w_i I_y(q_i)I_t(q_i) \end{bmatrix} \tag{5.14}$$

These weights are set to be as a Gaussian function of distance between neighbouring pixels and central pixel. In order for equation $A^T AV = A^T$b to be solvable, A^TA should be invertible, or eigen values of A^TA should satisfy $\lambda_1 \geq \lambda_2 > 0$. To avoid noise issue, usually λ_2 is required not to be too small. Also, if λ_1/λ_2 is too large, this means the central pixel is on an edge and this method suffers from the aperture problem. So, for this method to work properly, the condition is that, λ_1 and λ_2 should be large enough and have similar magnitude. This condition also holds true for Corner detection. This observation shows that one can easily tell which pixel is suitable for Lucas–Kanade method to work on by inspecting a single image. One main assumption for this method is that the motion is small (e.g. less than 1 pixel between two images). If the motion is large and violates this assumption then the problem can be solved by reducing the resolution of images first and then applying the Lucas-Kanade method.

Algorithm for Motion Detection

Step 1: Input video and detect Harris corners on which Lucas Kanade method is to be applied and then load its two consecutive frames.

Step 2: Resize both the frames entered in step1 by first doubling their precision and then reducing their size to half.

Step 3: Apply Lucas Kanade method on both the frames by finding I_x, I_y and I_t for every point by using ”convolution2” function.

Step 4: Now get A in matrix form such that A= $[I_x I_y]$.

Step 5: Now velocity vector is calculated i.e., V= $(A^T A)^{-1} A^T b$, where b = -I_t.

Step 6: These velocity vectors calculated in step above are downsized.

Step 7: Velocity vectors are plotted on second frame by using quiver plot starting position from change in first image and ending on second.

5.2.3 Tracking using Branching Particle Filter

A new class of Bayesian Filters [110, 111] called the Branching Particle Filters has been deployed in this chapter for object tracking. The Branching Particle Filters are designed to reduce variances but with different updating schemes. For the Branching Particle Filters, the updating is via branching in small time steps. Precisely, at each time step, each existing particle will die or give birth to a random number of offspring proportional to the weight. Particles that stay on the right tract (represented by heavy weights) are explored more thoroughly while particles with unlikely trajectories/positions (represented by little weights) are not carried forward uselessly.

A generic Particle Filter is studied by variation of the resampling process. In this process, the weakest sample is eliminated instead of the multiplication of the fittest sample. The genetic algorithm is used for reproducing from the fittest parents. This technique is used for replacing negligible weights.

In this technique, a partial resampling process occurs in place of full resampling for long time duration and a proper ratio of weight is allowed. In another case, when no resampling occurs, Weighted Particle Filters are used and weights may fluctuate accordingly. Similarly, in other extreme conditions, Branching Filters are used as an alternative to Particle Filters when full resampling occurs. This is the integrated method for residual resampling or combined PF. A changing rate of resampling occurs that would provide the effectiveness of compatibility between weighted variance increases & resampling noise whenever Branching Filters are used. The uniform random variables $\{U_n^K\}$ described by Branching Particle Filters and the branching variables $\{\rho_n^K\}$ are obtained from BPF Filters. These variables are separated in two different manners.

The Markov process is given as $\{S_n^N, n = 0, 1, ...\}$, It approximates the un-normalized filter $\{\sigma_n, n = 0, 1, ...\}$ in terms of the calculations as given below:

$\{X_O^k\}_{k=1}^N$ are independent samples of $\pi_O, N_O := N, N_n := 0, L_O^k := 1$ for k:=1,...... N and all n ∈ N

1 Weight by observation:

$$\hat{L}_{n+1}^K = \alpha_{n+1}(X_n^k)L_n^k \qquad for \qquad k := 1, 2, ...N_n \tag{5.15}$$

2 Evolve Independently:

$$Q^Y(\hat{X}_{n+1}^k \in \Gamma_k \forall k | F_n^X \vee F_{n+1}^U) = \Pi_{k=1}^{N_n} K(X_n^k, \Gamma_k) \forall \Gamma_k \tag{5.16}$$

3 Estimate:

$$\alpha_{n+1} \quad by; \quad S_{n+1}^N = \frac{1}{N}\sum_{k=1}^{N_n} \hat{L}_{n+1}^K \delta \hat{X}_{n+1}^K \quad and \quad \pi_{n+1}(f) \quad by \frac{S_{n+1}^N(f)}{S_{n+1}^N(1)} \tag{5.17}$$

4 Average Weight

$$A_{n+1} = S_{n+1}^N(1) \tag{5.18}$$

Repeat (5.17 to 5.18) : for k =1,2...... do

5 Resampling Case:

$$\hat{L}_{n+1}^K \in (\alpha_n An + 1, b_n An + 1) \qquad then \tag{5.19}$$

a Offspring Number

$$N_{n+1}^k = \left[\frac{\hat{L}_{n+1}^K}{A_{n+1}}\right] + \rho_{n+1}^k \quad with \quad \rho_{n+1}^k a \left(\frac{\hat{L}_{n+1}^K}{A_{n+1}} - \left[\frac{\hat{L}_{n+1}^K}{A_{n+1}}\right]\right) - Bernoulli \tag{5.20}$$

b Resample:

$$\hat{L}_{n+1}^{N_{n+1+j}} = A_{n+1}, \quad X_{n+}^{N_{n+1+j}} = \hat{X}_{n+1}^k \quad for \quad j = 1, \ldots\ldots, N_{n+1}^k \tag{5.21}$$

c Add Offspring Number:

$$N_{n+1} = N_{n+1} + N_{n+1}^{k} \tag{5.22}$$

6 Non-resample Case: If

$$\hat{L}_{n+1}^{K} \in (\alpha_n A_{n+1}, b_n A_{n+1}) \quad then \quad N_{n+1} = N_{(n+1)} + 1, \quad L_{n+1}^{N_(n+1)} = \hat{L}_{n+1}^{K}, X_{n+1}^{N_{n+1}} = \hat{X}_{n+1}^{k} \tag{5.23}$$

To avoid excess noise the estimated samples are extracted before the resampling process. In an unbiased manner new number of weights $L_{n+1}^{N_{n+1}}$ and particles N_{n+1} are determined by the step 5 & 6 steps of the algorithm. In this algorithm, splitting particles is determined by step 5 & for the kth particle, the prior weight $\hat{L}_{n+1}^{K}$ becomes extreme and beyond this residual style, branching is done. Since parents are at the same locations for having the average weights become zero in the given conditions. In step 6, run the weighted particles as there is no extreme condition for prior weight L_{Kn+1}. The extreme condition is determined in this class of algorithms for flexibility.

The proposed technique is used for duplicating and killing unlikely particles without biasing the particles in steps 3 to 5 but the total mass of particles and expected number of particles remain constant.

5.3 SIMULATION RESULTS OF BPF TRACKING

Different video datasets are used in this chapter for demonstrating the proposed tracking scheme effectively and robustly. Figure 5.4 gives the 108th frame out of the 110 frames of the video sequence called as KTH dataset (simple walk).

Figure 5.4: KTH Dataset, Simple Walk

The video sequence is that of a man walking at a constant speed in a single direction, with a static background. The results obtained after applying BPF to this dataset are shown in Figure 5.5.

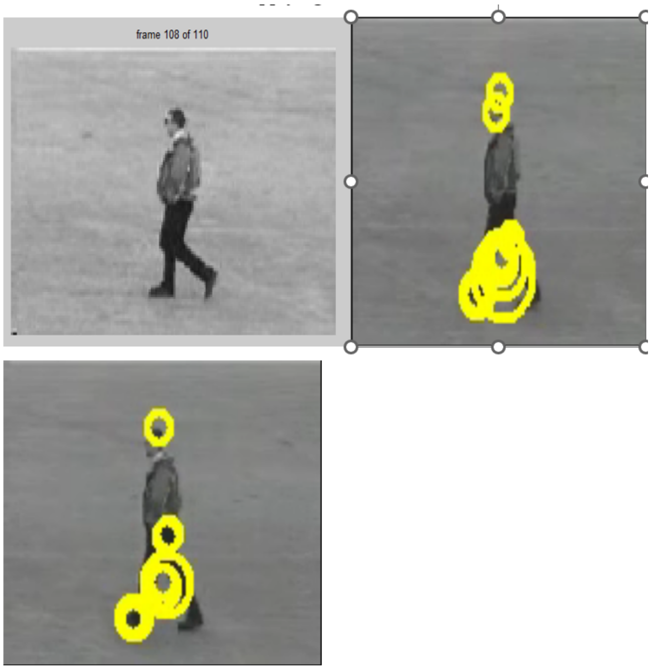


Figure 5.5: Tracking Performed on KTH Simple Walk Using BPF

Figure 5.6 depicts the frame of the video sequence KTH dataset (random walk) in which a man is walking in a dynamic or changing background. At different instances in the video, the person is occluded once by another person and once by a pole.

Figure 5.6: KTH Dataset, Complex Walk

The performance of the proposed Branching Particle Filter is evaluated by applying BPF to this dataset. The results so obtained are shown below in Figure 5.7. The results obtained clearly show that even though in the first frame the target person is occluded by a pole and in the second, he is occluded by another person yet the BPF does not lose track. It continues to track the correct target efficiently. It is now clear from these results that BPFs perform well in video scenes with dynamic background and also where the target object suffers occlusions. Hence, BPFs can be successfully applied to a high-dimensional system such as a crowded scene.

Figure 5.7: Tracking Performed on KTH Complex Walk using BPF

To prove this, the BPF is now applied to a third dataset called the UMN dataset (outdoor scene). The video consists of a crowd of people walking randomly outdoors. At different instances of time, the people are partially or fully occluded by each other. Figure 5.8 shows the original UMN dataset at different instances i.e. different frame numbers.

Figure 5.8: UMN Dataset, Outdoor Scene

The results obtained after applying the BPF to the UMN dataset are shown in Figure 5.9. As is evident from the figure, the Branching Particle Filter is successfully able to track each and every person in the crowded video scene where there is high probability of one person being occluded by another even at the time of panic or abnormal behavior.

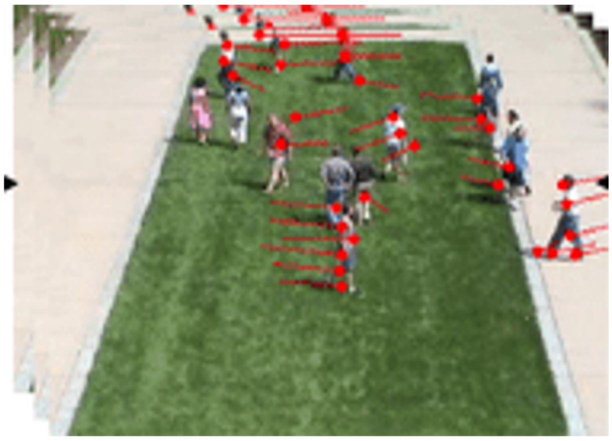


Figure 5.9: Tracking Performed on UMN Outdoor Scene using BPF

Comparison of Tracking Performance of the Modified Kalman Filter, Modified Particle Filter and Branching Particle Filter

The following parameters are chosen to evaluate and compare the tracking performance of the three filters namely the Modified Kalman Filter, the Modified Particle Filter and the Branching Particle Filter:

Detection Rate

This parameter is used for the measurement of the percentage of true targets detected. It may be defined as the ratio of the number of outliers detected to the total number of outliers in data.

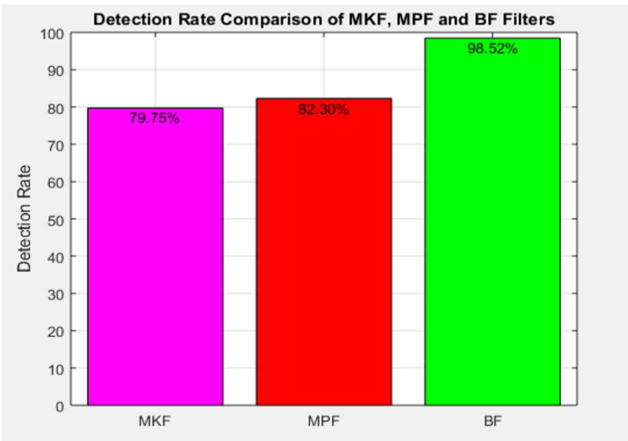


Figure 5.10: Comparison of Detection Rate of Modified Kalman Filter, Modified Particle Filter, and Branching Filter

The bar graph in Figure 5.10 compares the detection rate achieved by the proposed Branching Filter with the Modified Kalman and Modified Particle Filters as obtained in Chapter 3 and 4. The Kalman filter can detect anomalies at the rate of 80% approximately. The Particle Filter can achieve a better detection rate of 82% as compared to the traditional Kalman Filters. The Branching Particle Filters outperform the two filters by achieving a detection rate as high as 98%. It means Branching Filters are more effective as compared to other tracking methods for anomaly detection.

Accuracy

Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations.

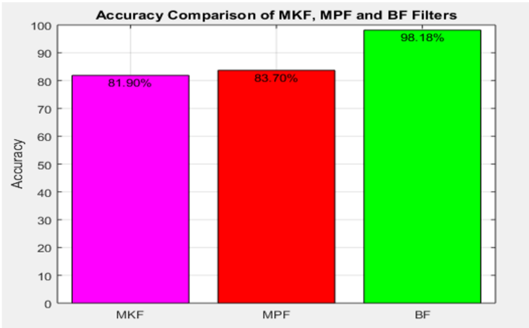


Figure 5.11: Comparison of Accuracy of Modified Kalman Filter, Modified Particle Filter, and Branching Particle Filter

The bar graph in Figure 5.11 compares the accuracy of the three filters under consideration. Whereas, the Branching Filter achieves accuracy as high as 98 %, the other two filters namely, Kalman and Particle Filters can accurately detect anomalous activities at 82% and 84% respectively.

RMSE

RMSE measures how much error there is between two data sets. In other words, it compares a predicted value and an observed or known value.

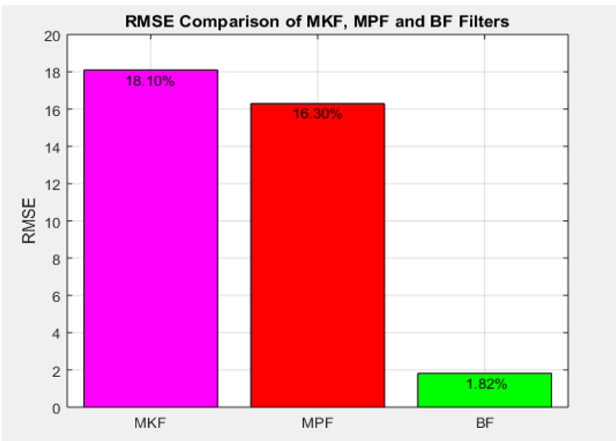


Figure 5.12: Comparison of Root Mean Square Error of Modified Kalman Filter, Modified Particle Filter, and Branching Particle Filter

The bar graph in Figure 5.12 compares the RMSE of the three filters. The graph clearly shows that the proposed Branching Filter outperforms the other two conventional filters as the RMSE obtained is as low as 0.01818. The RMSE obtained in Kalman and Particle Filters is 0.18 and 0.16 respectively.

The parameters compared in the above three bar graphs for the three filters under study are denoted in tabular form in Table 5.1

Table 5.1: Comparative Analysis of Kalman Filter, Particle Filter & Branching Filter

Sr. No.	Parameter	Modified Kalman Filter	Modified Particle Filter	Branching Filter (proposed)
1.	Detection Rate	79.75 %	82.30 %	98.52 %
2.	Accuracy	81.90 %	83.7 %	98.18 %
3.	RMSE	18.10 %	16.30 %	1.82 20%

5.4 ANOMALY DETECTION USING BRANCHING PARTICLE FILTER

Anomaly detection in crowd scenes is very important because of more concern for people's safety in a public places. As discussed in previous sections, the BPF had been applied to three different datasets, namely the KTH dataset (simple walk), the KTH dataset (complex walk), and the UMN dataset (outdoor scene). The simulation results so obtained shows the successful tracking of target objects under all three different scenarios using BPF. Since the BPF overcomes all the limitations of Kalman and Particle Filters in terms of tracking, the work is now proceeded further to detect anomalous activity in the dataset using BPF.

Figure 5.13 gives the framework for detecting anomalous activity in the UMN dataset outdoor scene using a Branching Particle Filter. Here an input training dataset is first taken as input. Then it is divided into frames. Feature points are then extracted using the Harris Corner Detector. The optical flow technique is applied at these corner points to extract motion vectors. Tracking of these motion vectors is performed using a Branching Particle Filter. This helps to track similar motion vectors. Similar motion vectors are then clustered using K means clustering technique [180] and the standard deviation value is calculated for the cluster which is set as the threshold value. Here the threshold value comes out to be 5. Now the test video is taken as input and similar steps are performed on the test data too. Motion patterns in the test video are tracked using a branching particle filter and they are compared with the cluster of similar motion patterns obtained from the training video.

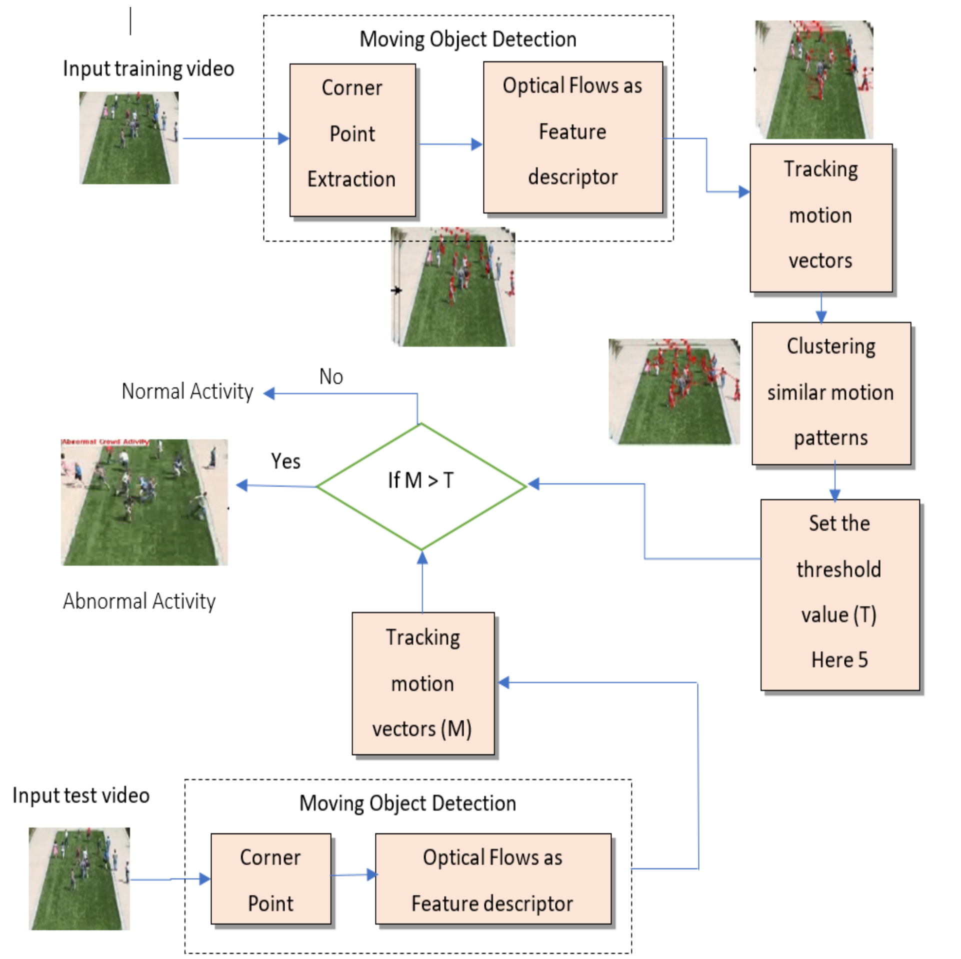


Figure 5.13: Anomaly Detection on UMN dataset using Branching Particle Filter

If the deviation of the motion vector from the cluster of similar motion vectors exceeds the threshold value, then the activity is considered to be anomalous otherwise the activity is considered normal.

5.4.1 Motion Descriptor

This research presents an approach to automatically detect abnormal behavior in a crowd scene. For this purpose, instead of tracking every person, Harris corners are extracted as feature points to represent moving objects and tracked by optical flow technique to generate motion vectors which are used to describe motion. This will decrease computation costs largely and retain rich motion information. Next, the whole frame is divided into small blocks and the motion pattern in each block is encoded by the distribution of motion vectors in it. Similar motion patterns are clustered into the pattern model in an unsupervised way and then the motion patterns are classified into normal or abnormal groups according to the deviation between the motion pattern and the trained model. The results of abnormal event detection in the real video demonstrates the effectiveness of the approach.

5.4.2 Cluster Motion Patterns

To generate a pattern model, similar motion patterns are clustered in an unsupervised way. In a crowd scene it cannot be known, how many kinds of movement there will be, hence, the number of clusters is not known. Therefore, an online cluster method is deployed which does not need the number of clusters first. The initial motion pattern is taken as the first pattern model and the deviation between all-new coming motion patterns and built models is then calculated. If the smallest deviation is greater than the threshold, this motion pattern can be treated as a new pattern model, if not, then it is considered that this motion pattern belongs to one pattern model which has the smallest deviation between them.

5.4.3 Measurement of Similarity

We define deviation measure as follows:

$$D(i,j) = \lambda_1(\mu_{vi} - \mu_{vj}) + \frac{\lambda_1}{\lambda_2}(\sigma_{vi}^2 - \sigma_{vj}^2) + \lambda_3(\mu_{ri} - \mu_{rj}) + \frac{\lambda_3}{\lambda_2}(\sigma_{ri}^2 - \sigma_{rj}^2) \tag{5.24}$$

where, D represents deviation, i and j represent the actual values and the measured values respectively. λ_1 and λ_3 are the weights of mean velocity μ_v and mean direction μ_r respectively and λ_2 is the weight of variance σ^2. These parameters are mainly dependent on what kind of behavior the user wants to detect. If the concern is more about velocity, λ_1 will be greater than λ_2 otherwise λ_2 will be greater. λ_1 and λ_2 will be similar if the concern is both with velocity and direction. λ_3 is used to keep variance weight being one-third of the mean weight. After the models have been trained, then in the test stage, motion patterns are classified into a normal or abnormal group according to the deviation between the coming motion pattern and models. If this deviation is greater than the threshold, this motion pattern is considered abnormal.

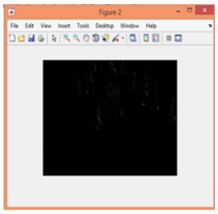
5.5 SIMULATION RESULTS OF BPF ANOMALY DETECTION

The simulation results obtained after applying Branching Particle Filters for tracking and utilizing motion vectors for deviation detection are shown below.

Figure 5.14a, 5.14b and 5.14c present the frame sequence of the actual dataset, the binary image representation and the average motion vector scores respectively. Since the average motion vector score lies below the threshold value except for one false detection, the frame is considered to be normal.

Similarly Figure 5.14d, 5.14e and 5.14f also depict one of the original frame

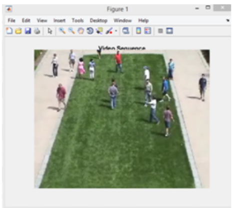
(a) UMN Original FS1



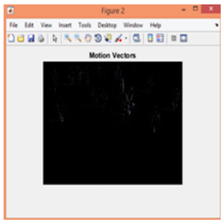
(b) Binary Image Representation of FS1

```
Command Window
New to MATLAB? See resources for Getting Started.
Average Motion Vector Score: 3.04
Average Motion Vector Score: 4.46
Average Motion Vector Score: 3.22
Average Motion Vector Score: 2.17
Average Motion Vector Score: 2.62
Average Motion Vector Score: 2.36
```

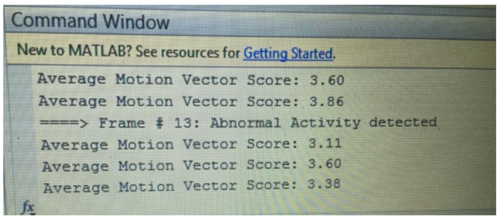
(c) Average Motion Vector Score for FS1



(d) UMN Original FS2

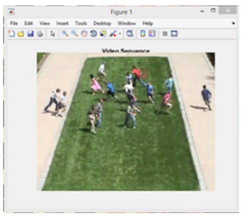


(e) Binary Image Representation of FS2

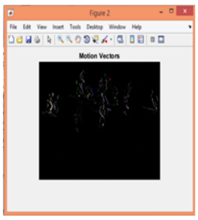


```
Command Window
New to MATLAB? See resources for Getting Started.
Average Motion Vector Score: 3.60
Average Motion Vector Score: 3.86
====> Frame # 13: Abnormal Activity detected
Average Motion Vector Score: 3.11
Average Motion Vector Score: 3.60
Average Motion Vector Score: 3.38
```

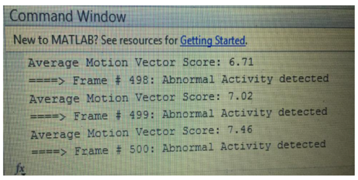
(f) Average Motion Vector Score for FS2



(g) Abnormal Activity Detected

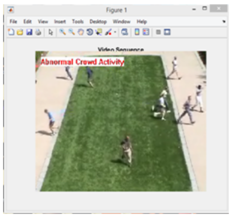


(h) Binary Image Representation

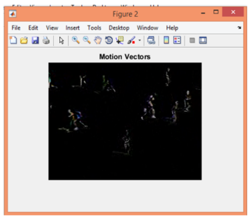


```
Command Window
New to MATLAB? See resources for Getting Started.
Average Motion Vector Score: 6.71
====> Frame # 498: Abnormal Activity detected
Average Motion Vector Score: 7.02
====> Frame # 499: Abnormal Activity detected
Average Motion Vector Score: 7.46
====> Frame # 500: Abnormal Activity detected
```

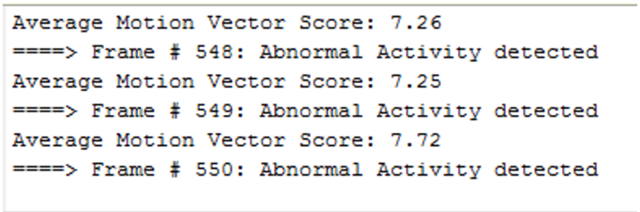
(i) Motion Vector Score for Abnormal Activity



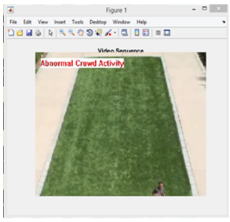
(j) Abnormal Activity Detected



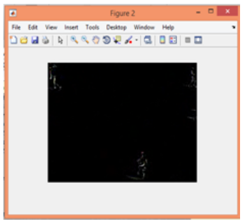
(k) Binary Image Representation



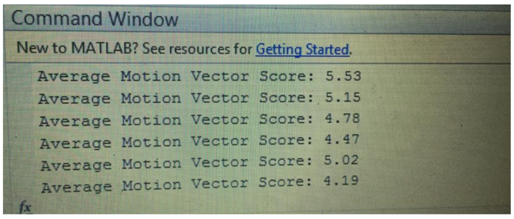
(l) Motion Vector Score for Abnormal Activity



(m) Abnormal Activity Detected



(n) Binary Image Representation



(o) Motion Vector Score for Abnormal Activity

Figure 5.14: Simulation Results of Anomaly Detection using BPF

sequences, its binary image and the average motion vector scores respectively. Again since the scores lie below the threshold, the frame is considered to have no abnormality.

Figure 5.14g shows that the people in the original frame sequence suddenly begin to run in random directions indicating some abnormal activity. As observed in Figure 5.14i, due to this sudden change in the kinetic velocity and direction of the motion of people, the motion vector score as calculated by the optical flow technique increases beyond the threshold limit, thus indicating an anomalous activity.

Similarly the deviation of the average motion vector score from the threshold value in Figure 5.14l indicates anomalous activity which conforms to the video sequence shown in Figure 5.14j, where the people can be seen running suddenly in all directions.

Figure 5.14m shows the frame sequence in which the people after running in all directions finally scatter and disappear out of the frame. Since no motion is detected in the frame, the average motion vector score remains below the threshold value indicating normal activity.

5.6 COMPARATIVE ANALYSIS OF BRANCHING FILTER

The simulation results obtained after applying Branching Particle Filter to various video datasets indicate that BPF gives excellent tracking and anomaly detection results in linear, nonlinear and high-dimensional crowd scenes, thereby overcoming the limitations of Kalman and Particle Filters. To validate the simulation results, a comparison of the three filters has been accomplished by taking into account various performance metrics. The parameters of BPF have also been compared with those given by previous researchers to prove the improved performance of the proposed work.

Table 5.2 gives the values of the various performance parameters for the Branching Particle Filter. The elapsed time that is the time taken by the algorithm to start giving its predictions once the code is run is 0.014653 seconds. This indicates that the algorithm works very fast and gives results almost immediately that is in run time or real-time.

Table 5.2: Performance Metrics for Evaluating the Performance of BPF

Sr. No.	Parameter	Score
1.	Total True Positives	79/80
2.	False Negative	1/80
3.	Total True Negatives	510/520
4.	Total False Positives	10/520
5.	Precision	88.76 %
6.	Recall	98.75 %
7.	F score	0.962
8.	RMSE	1.818 %
9.	Accuracy	98.182 %
10.	Detection Rate	98.52 %
11.	Elapsed Time	0.014653 seconds

The value of True Positives comes out to be 79 out of 80 which means that if the anomaly was present in the frame 80 times then out of 80, the filter was able to correctly predict the presence of anomaly 79 times and only one time did the algorithm missed out on detection of anomalous activity as given by 1/80 false negative score.

The true negatives score comes out to be 510/520 which tells us that out of the 520 frames in which there was no anomalous activity, the algorithm also reported correctly 510 times that there is no anomaly. Only 10 times false detections were made.

The proposed Branching Filter gives excellent precision and recall values of 88% and 98% respectively. The F score is 0.962 which is very close to its ideal value of 1. The RMSE is also as low as 1.818. The algorithm is accurate up to almost 98% and the detection rate is also very good at the value of 98.52%.

All these parameters indicate that the Branching Particle Filters give excellent results in nonlinear and high-dimensional crowd scenes.

Detailed interpretation of the above-mentioned parameters has been done with help of graphs and curves in the section below.

Precision vs Number of Particles: The graph in Figure 5.15 shows that precision of 89% is achieved at approximately 440 particles.

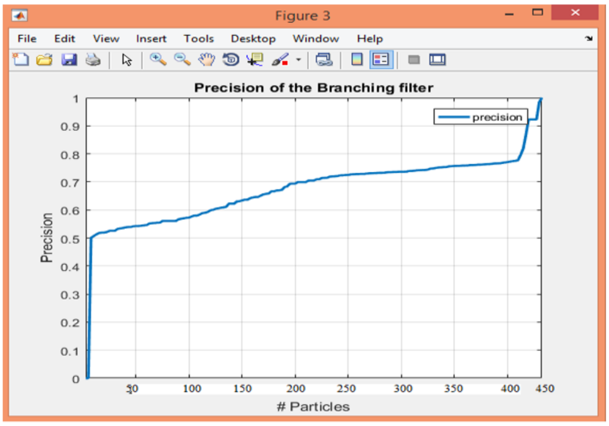


Figure 5.15: Precision vs. Number of Particles

Recall vs Number of Particles: The graph between Recall and number of particles in Figure 5.16 shows that a recall of approximately 99% is achieved at 450 particles.

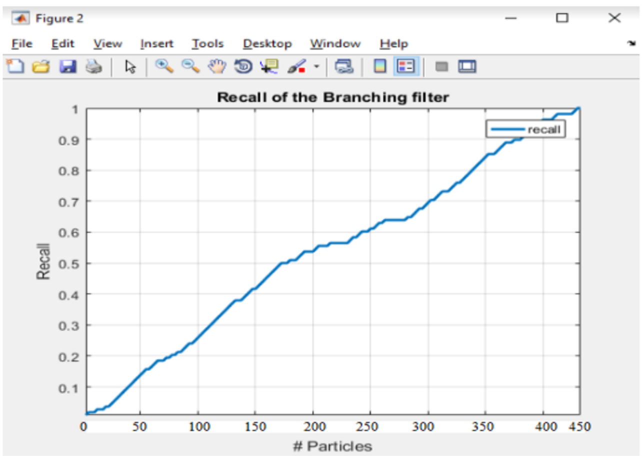


Figure 5.16: Recall vs. Number of Particles

Precision and Recall can be interpreted as follows:

If the algorithm is asked about the numbering of the 5 frames that contain the anomaly. If out of the 7 results returned by the algorithm, 5 of the frames that the algorithm mentioned are anomalous whereas 2 frames were incorrectly identified. So, it can be said that even though the recall of the algorithm was 100% (5/5), the precision was 71.4% (5/7).

F Score- The F Score is the harmonic mean of the precision and recall. It reaches its best value at 1 (perfect precision and recall).

The graph in Figure 5.17 below plots the F score with respect to the number of particles. The curve shows that as the number of particles increases, the F Score also increases. After 500 Particle iterations, the F Score becomes maximum and stable. The F score of the video is achieved up to 0.962. This particle number is quite less as compared to approximately 1000 particles required by particle filters.

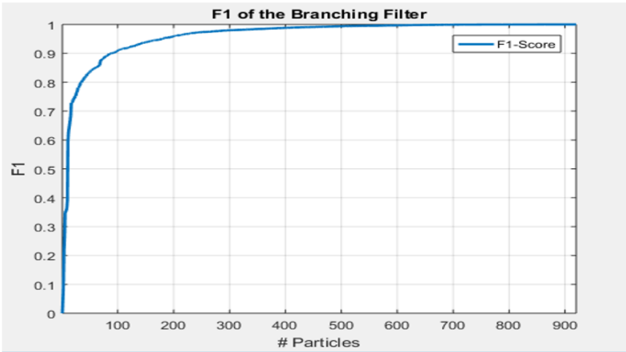


Figure 5.17: F Score Curve

Thus, it can be concluded that Branching Particle Filters give excellent results at a much lesser number of particles (approximately half the number of particles required by particle filters). This remarkably reduces the computational time and complexity of the algorithm.

ROC and AUROC

Receiver Operator Characteristic Curve or the ROC curve is a probability curve plotted between True Positive Rate and False Positive Rate where TPR and FPR are given by the formula:

$$\text{True Positive Rate} = \frac{TP}{TP+FN}$$
$$\text{False Positive Rate} = \frac{FP}{FP+TN}$$

Figure 5.18 gives the area under the ROC curve. AUROC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. The higher the AUC, the better the model is at predicting and distinguishing between normal and abnormal behavior.

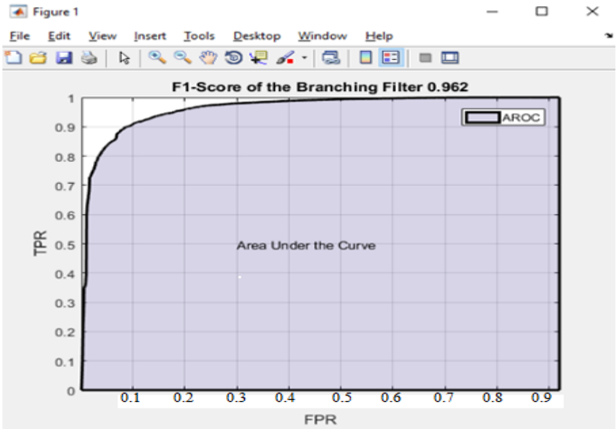


Figure 5.18: ROC and AUROC

Root Mean Square Error- The graph in Figure 5.19 is plotted between RMSE and the number of iterations. The curve tells that at 100 iterations, the RMSE is approximately 1.818 % i.e., 0.01818. This is much less as compared to RMSE values obtained by the Particle Filter and Kalman Filter.

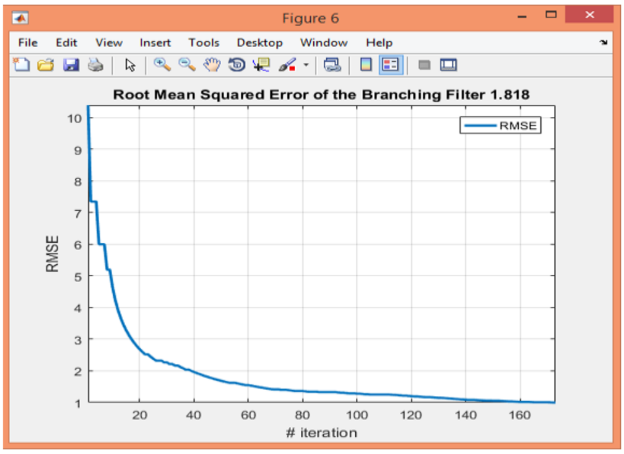


Figure 5.19: RMSE vs. Number of Iterations

The results obtained from the proposed work have been validated by comparison with other authors [2] in Table 5.3. In 2017 Haidar Sharif et al. [3] proposed the Eigen vector approach to detect flows and events in crowd videos. The recall, precision and Area Under Curve obtained by his method for the UMN dataset are 80.18%, 83.39% and 0.9783 respectively compared to 98.75%, 88.76% and 0.962 obtained from the proposed technique.

This indicates that the Branching Filter outperforms the previous techniques [2, 3] for anomaly detection in crowd videos by giving much better values of precision and recall at an almost equal value of the area under the curve.

Table 5.3: Performance Comparison of Proposed work with Previous Literature [2, 3] for UMN Dataset

Method	AUC	Accuracy	Precision	Recall	F Score
Shear Force [2]	0.929	89.7%	0.887	0.912	0.899
HOG/HOF [2]	0.875	84.5%	0.832	0.878	0.854
HOG 3D[2]	0.771	73.3%	0.812	0.793	0.802
Cuboid 3D [2]	0.801	73.8%	0.837	0.810	0.823
Local Motion [2]	0.789	76.2%	-	-	-
Tracklets [2]	0.919	85.2%	-	-	-
Social Force [2]	0.912	80.2%	0.792	0.891	0.838
Pure Optical Flow[2]	-	-	0.803	0.822	0.812
Haidar Sharif [3]	0.9783	92.6%	0.8339	0.8018	0.817
Proposed	0.962	98.1%	0.8876	0.9875	0.934

Figure 5.20 gives the bar graph representation of Table 5.3. It gives a comparative analysis of the Branching Particle Filter with the previous literature [2, 3] in terms of accuracy and area under the curve. Figure 5.21 further compares the performance of the Branching Particle Filter with the previous techniques [2, 3] in terms of precision, recall and F score.

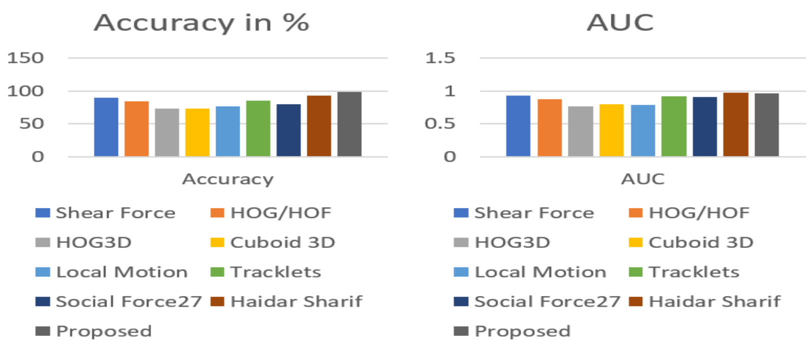


Figure 5.20: Comparative Analysis of Branching Particle Filter in terms of Accuracy and AUC

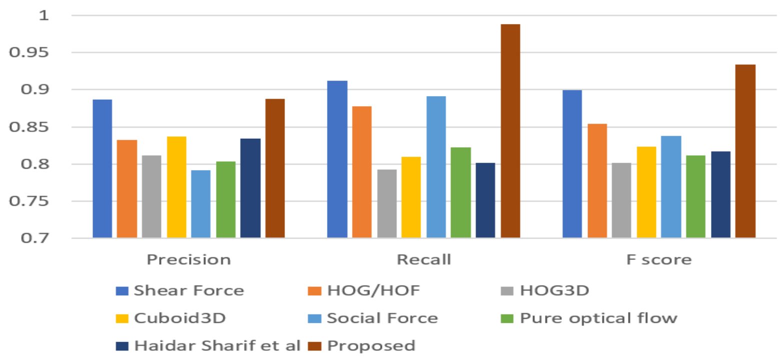


Figure 5.21: Comparative Analysis of Branching Particle Filter in terms of Precision, Recall, and F Score

5.7 RESULTS AND DISCUSSION

This chapter deploys the Branching Particle Filter to track the object of interest. The performance of the filter is evaluated by applying the filter to various datasets. As observed in Figures 5.5 and 5.7, the Branching Filter is successfully able to track the target person in both cases, first, when the person is moving in the static background and is not occluded and in the second case, when the object is moving in the dynamic background and is occluded by other objects. Once it is proved that Branching Filters can successfully track the object of interest even when it is completely occluded, the work is extended to evaluate the performance of the filter on a high-dimensional dataset. Figure 5.8 shows the UMN dataset, an outdoor scene that depicts a crowd of people walking randomly in all directions. The Branching Filter when applied to this dataset is successfully able to track all the people throughout the video even when the people are partially or completely occluded by each other at certain instances as shown in Figure 5.9.

This simulation result is validated by comparing three performance parameters namely the detection rate, accuracy and RMSE of the modified Kalman and Particle Filters with the proposed Branching Filter on the UMN dataset as shown in Table 5.2. For all three parameters, it has been observed that the Branching Filter outperforms the other two filters. This is because the Kalman Filter assumes a linear Gaussian environment so its detection rate and accuracy are poor on the UMN dataset which depicts a nonlinear non-Gaussian environment. The Particle Filter also gives poor parameter values because in the UMN dataset there is a crowd of people in which people are occluded by each other. The Branching Filter overcomes this issue because the particles affected due to occlusion are diversified in the search space by the BPF to improve their placement. This helps in maintaining the tracker's efficiency under full or partial occlusion. Moreover, in the case of Particle Filters, at each resampling step, noise particles are also resampled thereby giving a high RMSE value. In BPF, particles are branched individually rather than resampled collectively and branching only occurs when there is sufficient need. With this limited branching, noise particles are not resampled and RMSE decreases.

Table 5.3 further compares the performance of the BPF with the other state-of-the-art techniques [2, 3] based on performance parameters namely accuracy, precision, recall, F Score and area under the curve. Once again it is observed that the BPF outperforms the technique proposed by Haidar Sharif et al. [3] This is because the Eigen vector approach of Haider Sharif et al. [3] deals only with multiple separate instances of single flow occurring in videos with temporal overlap. It is not able to deal with multiple separate instances of single flow occurring in videos with no temporal overlap i.e. the crowd of people walking in random directions simultaneously. This is because for flow detection in a particular direction a specific threshold value is decided upon.

If the crowd will move in any and every direction at the time of catastrophe (considered an abnormal activity), then the eigen vector approach is not able to decide that it needs to reevaluate different threshold values every time. Moreover, the approach of Haider Sharif et al. [3] does not work in real-time. It evaluated the parameters in an offline fashion. The BPF overcomes all these issues. It can track and detect the abnormal activity of the crowd and also works in real-time.

5.8 CONCLUSIONS

This study proposes a novel framework to overcome the problems in existing video anomaly detection techniques, such as:

1. False anomaly detection in nonlinear and non-Gaussian environments.
2. False anomaly detection in crowded scenes where the object of interest might be partially or fully occluded.

The proposed technique is based on anomaly detection using object tracking. The Branching Filtering Technique provides the best solution to achieve maximum accuracy. Based upon the tests performed and hypothetical outcomes, it is proposed that Branching Particle Filters:

a. Detect abnormal activity in a crowded scene with dynamic background and partial or full occlusions.

b. Avoid the particle spread problems of the weighted-particle filter.

c. Runs significantly faster than particle filters on tracking and other Bayesian models.

The Branching Particle Filter is effective when extended to estimate the conditional distribution for multiple, varying and unknown targets, provided, the number of targets is small. If the number of targets is greatly increased i.e. in very high dimensional datasets, the Branching Filters do not give very good results. This is due to possible instability of the number of particles as well as the computational consequences of this instability. This issue arises from the particle number drift property explained by Kouritzin et al. [111]

Chapter 6

CONCLUSIONS AND FUTURE SCOPE

6.1 CONCLUSIONS

Detecting anomalies or abnormal events in surveillance videos is gaining high demand and importance from the point of view of domestic, commercial and national security. Anomaly detection techniques in videos can be broadly classified into supervised, semi-supervised and unsupervised techniques. Most of the existing techniques are supervised or semi supervised i.e. they require a large database of images that have been labeled as ”normal” which is then used for training a classifier. Considering the vast applications of surveillance cameras and the enormous amount and variety of data collected by them, it seems almost impossible to label a database as normal or abnormal. The goal of the current study is to overcome the limitations of the supervised and semi-supervised techniques by developing a practical anomaly detection system which is unsupervised and runs in real time, so as to timely signal an activity that deviates from the normal patterns.

This research work uses Bayesian Filters for unsupervised anomaly detection in videos. The concept behind using Bayes Filters for anomaly detection is that these filters predict the likelihood or occurrence of a particular measurement, given all previous measurements. The algorithm works by analyzing all the previous measurements so as to decide upon a threshold. This threshold will be the deciding factor to label events as normal or anomalous. Now the object of interest is tracked using these filters and their likely position or behavior is predicted. If the difference between the measured value and the predicted value lies within the threshold then the event or behavior is considered normal. If this difference exceeds the threshold then the activity is considered to be anomalous.

Three kinds of Bayesian Filters have been deployed in this study namely the Modified Kalman Filter, the Modified Particle Filter and the Branching Particle Filter. The first and simplest of all the Bayes' Filters deployed for this purpose is the Modified Kalman Filter.

Object tracking using Modified Kalman Filter is performed by predicting the object's position from the previous information and verifying the existence of the object at the predicted position. Kalman Filter uses two steps while estimating the unknown states:

i. Prediction step

ii. Updating step

The predict phase uses the state estimate from the previous timestep to produce an estimate of the state at the current timestep. In the update phase, the current prediction is combined with current observation information to refine the state estimate.

- Kalman Filter gives good tracking results when applied to datasets with linear and Gaussian distributed video data.

- Kalman Filters show poor tracking performance in environments which introduce nonlinear factors such as body moving at non-uniform speed or multipath propagation effects as well as crowded scenes. Also, if the movements are rapid and unpredictable (e.g. leaf on a tree during windy day), the Kalman filter is likely to fail. This is because Kalman Filter assumes that both the system and observation model equations are linear, which is not likely in many real-life situations. Also, it assumes that the state belief is Gaussian distributed.

- Kalman Filters also fail to give good tracking results if the object of interest is occluded by another object even though partially.

To overcome the limitation of Kalman Filter, Modified Particle Filter was then deployed to track the object of interest. Apart from the two steps, prediction and update in case of Kalman Filters, Particle Filters involve a third step called resampling.

i. Prediction

ii. Update

iii. Resampling

Resampling procedure eliminates the particles that become redundant and multiplies the ones that contribute most to the resulting approximation. The performance of Particle Filter was evaluated by applying it to the dataset on which the Kalman Filter had shown poor results. The Particle Filter showed excellent tracking results in such a scenario which is shown in Figure 4.5b. Once the Particle Filters showed good tracking results on dataset 2 the work was extended to evaluate the performance of the Particle Filter on a dataset which depicted high dimensional data such as a dense crowd of people at a fair or a historical monument where the object of interest might be partially of fully occluded by another object. Such a situation is depicted in dataset 3 in Figure 4.7a. On applying Particle Filters to track the object of interest in such a high dimensional dataset it was found that the Particle Filters failed to track the object of interest which was occluded by another object as seen in Figure 4.7b. Following conclusions are drawn based on the observations and findings:

- Particle Filters work extremely well in nonlinear non-Gaussian environment, thereby overcoming the limitations of Kalman Filters.
- With the Particle Filter approach, the system will be able to model the partially available information from samples which give a higher chance for the tracker to track the objects even in the case of partial occlusion.
- Particle Filters are successful in tracking objects that are occluded by other objects but only very partially. If the target object is completely or much occluded then Particle Filter fails to track occluded objects. Basic Particle Filters give poor results in tracking objects in crowded video sequences where there is a possibility of the target object to get completely occluded.

Since Particle Filters failed to track the object of interest in high dimensional dataset, the novel Branching Particle Filters were introduced to the field of object tracking and anomaly detection in videos. Branching Filters differ from the Particle Filters in the fact that unlike Particle Filters, the updating in Branching Filters is via branching in small time steps and also resampling of the particles is not done regularly in an unbiased manner. In BPFs resampling is done only when there is sufficient need. To evaluate the performance of BPF, the filter was applied to the KTH dataset for both simple and complex walk shown in Figures 5.4 and 5.6. The BPF showed excellent results under both the conditions which is depicted in Figures 5.5 and 5.7. Next the filter was applied to the UMN dataset which represents a high dimensional environment as shown in Figure 5.8. Since the Branching Filter outperformed the other two filters in tracking the object of interest in high dimensional dataset (Figure 5.9), the research work proceeds further to deploy the Branching Filter to detect abnormal crowd activity in the UMN dataset.

The Branching Filter gave better performance in detecting abnormal activity in the above given dataset as compared to the techniques used by previous researchers. The results obtained from simulation were validated by assessing the Branching Filters on various performance parameters and by comparing the value of these parameters with those obtained by past techniques. Following conclusions are drawn based on the observations and findings:

- The Branching Filter is able to successfully track the object of interest. This is because the particles affected due to occlusion are diversified in the search space by the BPF in order to improve their placement. This helps in maintaining the tracker's efficiency under full or partial occlusion. Thus, it can be concluded that Branching Filters detect abnormal activity in crowded scene with dynamic background under partial or full occlusions. Moreover, BPFs are stable with respect to the number of particles.

- The research work also concludes that with an accuracy of 98.18, the BPFs outperform the Kalman and Particle Filters that give an accuracy of 81.90 and 83.70 respectively.

- The above inference is also supported when the three filters are compared based on the values of RMSE. It was found that the RMSE was as low as 1.82 in comparison to 18.10 and 16.30 shown by the other two filters respectively.

- The rate at which the Branching Filters are able to detect the anomalies is also much better than the other two filters. The Kalman, Particle and Branching Filter give a detection rate of 79.75, 82.30 and 98.52 respectively.

- Lastly the proposed Branching Filter was also compared with the other two filters based on the parameter "elapsed time". It was found that BPFs with an elapsed time of just 0.014 seconds, run significantly faster on tracking than Particle Filters and other Bayesian models.

6.2 FUTURE SCOPE

The research work though endeavors to introduce a technique that gives better performance as compared to the techniques put forward by previous researchers yet during the course of study it was found that the proposed work suffered certain limitations that can in the future inspire researchers to bring improvements in the proposed work.

The Branching Particle Filter has been found to be effective when extended to estimate the conditional distribution of multi-target signals for unknown, varying, but only with small numbers of targets. When the number of targets to be tracked in a highly dense crowd is much high, the Branching Filters may tend to give poor results. This issue can be resolved by using some extended or advanced Branching Filters that are already available but they have yet not been applied to the field of video anomaly detection. Example of such filters are the Residual Filters or the Combined Filters [183]. Thus, future work may include the introduction of Residual and Combined Branching Filters in object tracking and anomaly detection for greater number of targets.

Also, the current study limits the use of Bayes' Filters to anomaly detection in surveillance videos. This section briefs how the scope or application of proposed technique can be extended to areas other than surveillance videos. Branching Filters can be used for mobile nodes tracking in Wireless Sensor Network [184] framework. Target tracking is one of the most fundamental applications in WSN in which sensor nodes constantly monitor and report the positions of moving objects. In surveillance applications, sensor nodes produce data almost in real-time while tracking the objects in a critical area or monitoring border activities. Owing to multiple paths, occlusions, and recalibration effects, the tracking process suffers from low stability and precision. To overcome these limitations, the Branching Filters may be introduced in the field of wireless sensor networks.

REFERENCES

[1] X. Zhang, S. Yang, J. Zhang, and W. Zhang, "Video anomaly detection and localization using motion-field shape description and homogeneity testing," *Pattern Recognition*, vol. 105, pp. 1–13, 2020.

[2] H. Su, H. Yang, S. Zheng, Y. Fan, and S. Wei, "The large-scale crowd behavior perception based on spatio-temporal viscous fluid field," *IEEE Transactions on Information Forensics and security*, vol. 8, no. 10, pp. 1575–1589, 2013.

[3] M. H. Sharif, "An eigenvalue approach to detect flows and events in crowd videos," *Journal of Circuits, Systems and Computers*, vol. 26, no. 07, pp. 1–50, 2017.

[4] V. Tsakanikas and T. Dagiuklas, "Video surveillance systems-current status and future trends," *Computers & Electrical Engineering*, vol. 70, pp. 736–753, 2018.

[5] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6479–6488.

[6] T. Xiang and S. Gong, "Video behavior profiling for anomaly detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 5, pp. 893–908, 2008.

[7] S. Ahmad, A. Lavin, S. Purdy, and Z. Agha, "Unsupervised real-time anomaly detection for streaming data," *Neurocomputing*, vol. 262, pp. 134–147, 2017.

[8] S. Agrawal and J. Agrawal, "Survey on anomaly detection using data mining techniques," *Procedia Computer Science*, vol. 60, pp. 708–713, 2015.

[9] R. Aarthi, J. Amudha, K. Boomika, and A. Varrier, "Detection of moving objects in surveillance video by integrating bottom-up approach with knowledge base," *Procedia Computer Science*, vol. 78, pp. 160–164, 2016.

[10] Q. Chen, Z. Song, J. Dong, Z. Huang, Y. Hua, and S. Yan, "Contextualizing object detection and classification," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 1, pp. 13–27, 2014.

[11] Z. Zhao, P. Zheng, S. Xu, and X. Wu, “Object detection with deep learning: A review,” *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[12] V. Kastrinaki, M. Zervakis, and K. Kalaitzakis, “A survey of video processing techniques for traffic applications,” *Image and vision computing*, vol. 21, no. 4, pp. 359–381, 2003.

[13] Q. Xie, O. Remil, Y. Guo, M. Wang, M. Wei, and J. Wang, “Object detection and tracking under occlusion for object-level RGB-D video segmentation,” *IEEE Transactions on Multimedia*, vol. 20, no. 3, pp. 580–592, 2017.

[14] G. Sreenu and S. Durai, “Intelligent video surveillance: a review through deep learning techniques for crowd analysis,” *Journal of Big Data*, vol. 6, no. 1, pp. 1–27, 2019.

[15] H. Izadinia, I. Saleemi, and M. Shah, “Multimodal analysis for identification and segmentation of moving-sounding objects,” *IEEE Transactions on Multimedia*, vol. 15, no. 2, pp. 378–390, 2012.

[16] A. Faro, D. Giordano, and C. Spampinato, “Adaptive background modeling integrated with luminosity sensors and occlusion processing for reliable vehicle detection,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1398–1412, 2011.

[17] K. Patel, “Motion detection and segmentation in H.264 compressed domain for video surveillance application,” *International Journal of Engineering Research and Technology (IJERT)*, vol. 3, no. 4, 2014.

[18] Y. Zhao, “Candidate smoke region segmentation of fire video based on rough set theory,” *Journal of Electrical and Computer Engineering*, vol. 2015, 2015.

[19] F. Xu, G. Gu, K. Ren, and W. Qian, “Motion segmentation by new three-view constraint from a moving camera,” *Mathematical Problems in Engineering*, vol. 2015, 2015.

[20] Y. Luo and Y. Guan, “Enhanced facial texture illumination normalization for face recognition,” *Applied optics*, vol. 54, no. 22, pp. 6887–6894, 2015.

[21] Z. Wu, F. Li, R. Sukthankar, and J. M. Rehg, “Robust video segment proposals with painless occlusion handling,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4194–4203.

[22] Y. Wan, X. Wang, and H. Hu, "Automatic moving object segmentation for freely moving cameras," *Mathematical Problems in Engineering*, vol. 2014, 2014.

[23] C. Zhang and H. Ai, "Video object segmentation by hierarchical localized classification of regions," in *The First Asian Conference on Pattern Recognition*. IEEE, 2011, pp. 244–248.

[24] M. Sun, J. Xiao, E. G. Lim, Y. Xie, and J. Feng, "Adaptive ROI generation for video object segmentation using reinforcement learning," *Pattern Recognition*, vol. 106, p. 107465, 2020.

[25] J. Jiang and X. Song, "An optimized higher order CRF for automated labeling and segmentation of video objects," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 3, pp. 506–516, 2015.

[26] S. Avinash Ramakanth and R. Venkatesh Babu, "Seamseg: Video object segmentation using patch seams," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 376–383.

[27] Z. Zhou, W. Luo, Q. Wang, J. Xing, and W. Hu, "Distractor-aware discrimination learning for online multiple object tracking," *Pattern Recognition*, vol. 107, p. 107512, 2020.

[28] M. Y. Yang and B. Rosenhahn, "Video segmentation with joint object and trajectory labeling," in *IEEE Winter Conference on Applications of Computer Vision*. IEEE, 2014, pp. 831–838.

[29] S. Chien, W. Chan, D. Cherng, and J. Chang, "Human object tracking algorithm with human color structure descriptor for video surveillance systems," in *2006 IEEE International Conference on Multimedia and Expo*. IEEE, 2006, pp. 2097–2100.

[30] P. Chen, H. Krim, and O. Mendoza, "Multiphase joint segmentation-registration and object tracking for layered images," *IEEE transactions on image processing*, vol. 19, no. 7, pp. 1706–1719, 2010.

[31] A. W. Min and K. G. Shin, "Robust tracking of small-scale mobile primary user in cognitive radio networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 4, pp. 778–788, 2012.

[32] H. Medeiros, G. Holguín, P. J. Shin, and J. Park, "A parallel histogram-based particle filter for object tracking on SIMD-based smart cameras," *Computer Vision and Image Understanding*, vol. 114, no. 11, pp. 1264–1272, 2010.

[33] T. Mélange, M. Nachtegael, S. Schulte, and E. E. Kerre, "A fuzzy filter for the removal of random impulse noise in image sequences," *Image and Vision Computing*, vol. 29, no. 6, pp. 407–419, 2011.

[34] Z. Cai and N. Vasconcelos, "Cascade R-CNN: high quality object detection and instance segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 5, pp. 1483–1498, 2019.

[35] J. Ding, Y. Huang, W. Liu, and K. Huang, "Severely blurred object tracking by learning deep image representations," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 2, pp. 319–331, 2015.

[36] T. Nawaz, F. Poiesi, and A. Cavallaro, "Measures of effective video tracking," *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 376–388, 2013.

[37] S. Zhang, X. Yu, Y. Sui, S. Zhao, and L. Zhang, "Object tracking with multi-view support vector machines," *IEEE Transactions on Multimedia*, vol. 17, no. 3, pp. 265–278, 2015.

[38] H. Sakaino, "Video-based tracking, learning, and recognition method for multiple moving objects," *IEEE transactions on circuits and systems for video technology*, vol. 23, no. 10, pp. 1661–1674, 2013.

[39] W. Li, Y. Hu, X. Fu, S. Lu, and D. Chen, "Cooperative positioning and tracking in disruption tolerant networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 2, pp. 382–391, 2014.

[40] L. Wang, T. Liu, G. Wang, K. L. Chan, and Q. Yang, "Video tracking using learned hierarchical features," *IEEE Transactions on Image Processing*, vol. 24, no. 4, pp. 1424–1435, 2015.

[41] X. Song, H. Zhao, J. Cui, X. Shao, R. Shibasaki, and H. Zha, "An online system for multiple interacting targets tracking: Fusion of laser and vision, tracking and learning," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 4, no. 1, pp. 1–21, 2013.

[42] Q. Wang, F. Chen, J. Yang, W. Xu, and M. H. Yang, "Transferring visual prior for online object tracking," *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3296–3305, 2012.

[43] X. Song, Q. Zhang, Y. Sekimoto, R. Shibasaki, N. J. Yuan, and X. Xie, "Prediction and simulation of human mobility following natural disasters," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 8, no. 2, pp. 1–23, 2016.

[44] S. Salti, A. Cavallaro, and L. Di Stefano, “Adaptive appearance modeling for video tracking: Survey and evaluation,” *IEEE Transactions on Image Processing*, vol. 21, no. 10, pp. 4334–4348, 2012.

[45] B. Deori and D. M. Thounaojam, “A survey on moving object tracking in video,” *International Journal on Information Theory (IJIT)*, vol. 3, no. 3, pp. 31–46, 2014.

[46] X. Li, W. Hu, C. Shen, Z. Zhang, A. Dick, and A. V. D. Hengel, “A survey of appearance models in visual object tracking,” *ACM transactions on Intelligent Systems and Technology (TIST)*, vol. 4, no. 4, pp. 1–48, 2013.

[47] J. J. Jadhav and Y. R. Patil, “Moving object detection for video surveillance system,” *International Journal of Computer Applications*, vol. 975, pp. 7–10, 2015.

[48] J. Sukanyathara and A. Kuriakose, “An optimized framework for detection and tracking of video objects in challenging backgrounds,” *The International Journal of Multimedia & Its Applications*, vol. 6, no. 4, pp. 27–38, 2014.

[49] L. Zhang, H. Wang, and R. Wu, “A high-capacity steganography scheme for JPEG 2000 baseline system,” *IEEE transactions on Image Processing*, vol. 18, no. 8, pp. 1797–1803, 2009.

[50] S. Bhattacharya, H. Idrees, I. Saleemi, S. Ali, and M. Shah, “Moving object detection and tracking in forward looking infra-red aerial imagery,” in *Machine vision beyond visible spectrum*. Springer, 2011, pp. 221–252.

[51] S. W. Sun, Y. C. F. Wang, F. Huang, and H. Y. M. Liao, “Moving foreground object detection via robust SIFT trajectories,” *Journal of Visual Communication and Image Representation*, vol. 24, no. 3, pp. 232–243, 2013.

[52] J. Han and Z. Pan, “Robust moving objects detection in dynamic scenes based on mean shift,” in *2008 International Conference on Cyberworlds*. IEEE, 2008, pp. 271–275.

[53] S. Vahora, N. Chauhan, and N. Prajapati, “A robust method for moving object detection using modified statistical mean method,” *International Journal of Advanced Information Technology*, vol. 2, no. 1, pp. 65–73, 2012.

[54] L. Tong, F. Dai, Y. Zhang, and J. Li, “Visual security evaluation for video encryption,” in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 835–838.

[55] F. Zhang, J. Yang, Y. Tai, and J. Tang, “Double nuclear norm-based matrix decomposition for occluded image recovery and background modeling,” *IEEE Transactions on Image Processing*, vol. 24, no. 6, pp. 1956–1966, 2015.

[56] S. Pudlewski and T. Melodia, “Compressive video streaming: Design and rate-energy-distortion analysis,” *IEEE Transactions on Multimedia*, vol. 15, no. 8, pp. 2072–2086, 2013.

[57] S. V. Kothiya and K. B. Mistree, “A review on real time object tracking in video sequences,” in *International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO), 2015*, 2015, pp. 1–4.

[58] Z. Q. Zhao, P. Zheng, S. t. Xu, and X. Wu, “Object detection with deep learning: A review,” *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[59] W. Hu, X. Zhou, W. Li, W. Luo, X. Zhang, and S. Maybank, “Active contour-based visual tracking by integrating colors, shapes, and motions,” *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1778–1792, 2012.

[60] D. Varas and F. Marques, “Region-based particle filter for video object segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 3470–3477.

[61] K. Q. Weinberger and L. K. Saul, “Distance metric learning for large margin nearest neighbor classification.” *Journal of machine learning research*, vol. 10, no. 2, pp. 207–244, 2009.

[62] M. J. Patel and B. Bhatt, “A comparative study of object tracking techniques,” *International Journal of Innovative Research in Science, Engineering and Technology*, vol. 4, no. 3, pp. 1361–1364, 2015.

[63] S. P. Coraluppi and C. A. Carthel, “Multiple-hypothesis tracking for targets producing multiple measurements,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 3, pp. 1485–1498, 2018.

[64] G. Welch and G. Bishop, “An introduction to the kalman filter,” in *SIGGRAPH 2001*. Chapel Hill, NC, USA, 1995.

[65] W. Zhang, G. Du, S. Zhang, and Y. Zhang, “Study of nonlinear filter methods: particle filter,” *Journal of systems Engineering and Electronics*, vol. 17, no. 1, pp. 1–5, 2006.

[66] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 25, no. 5, pp. 564–577, 2003.

[67] N. D. Thanh, W. Li, and P. Ogunbona, "An improved template matching method for object detection," in *Asian Conference on Computer Vision*. Springer, 2009, pp. 193–202.

[68] I. A. Iswanto and B. Li, "Visual object tracking based on mean-shift and particle-kalman filter," *Procedia computer science*, vol. 116, pp. 587–595, 2017.

[69] K. Huang, H. Jiang, and X. Y. Zhang, "Field support vector machines," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 1, no. 6, pp. 454–463, 2017.

[70] X. Cao, X. Jiang, X. Li, and P. Yan, "Correlation-based tracking of multiple targets with hierarchical layered structure," *IEEE transactions on cybernetics*, vol. 48, no. 1, pp. 90–102, 2016.

[71] M. Wu, H. Ling, N. Bi, S. Gao, Q. Hu, H. Sheng, and J. Yu, "Visual tracking with multiview trajectory prediction," *IEEE Transactions on Image Processing*, vol. 29, pp. 8355–8367, 2020.

[72] D. P. Chau, F. Bremond, and M. Thonnat, "Object tracking in videos: Approaches and issues," *HAL Id: hal-00815046 https://hal.inria.fr/hal-00815046*, 2013.

[73] S. Kanagamalliga and S. Vasuki, "Contour-based object tracking in video scenes through optical flow and gabor features," *Optik*, vol. 157, pp. 787–797, 2018.

[74] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 4, pp. 509–522, 2002.

[75] T. Kamp, T. Rolo, P. Vagovic, T. Baumbach, and A. Riedel, "Three-dimensional reconstructions come to life–interactive 3D PDF animations in functional morphology," *PLoS One*, vol. 9, no. 7, pp. 1–7, 2014.

[76] M. V. Afonso, J. C. Nascimento, and J. S. Marques, "Automatic estimation of multiple motion fields from video sequences using a region matching based approach," *IEEE Transactions on Multimedia*, vol. 16, no. 1, pp. 1–14, 2013.

[77] H. Shen, S. Li, C. Zhu, H. Chang, and J. Zhang, "Moving object detection in aerial video based on spatiotemporal saliency," *Chinese Journal of Aeronautics*, vol. 26, no. 5, pp. 1211–1217, 2013.

[78] P. Chiranjeevi and S. Sengupta, "Interval-valued model level fuzzy aggregation-based background subtraction," *IEEE transactions on cybernetics*, vol. 47, no. 9, pp. 2544–2555, 2016.

[79] A. G. Dopico, M. V. Correia, J. A. Santos, and L. M. Nunes, "Parallel computation of optical flow," in *International Conference Image Analysis and Recognition*. Springer, 2004, pp. 397–404.

[80] S. Colonnese, F. Cuomo, O. Damiano, V. Pascalis, and T. Melodia, "On the adoption of multiview video coding in wireless multimedia sensor networks," in *Wireless Advanced*. IEEE, 2011, pp. 218–223.

[81] L. Lin, D. Zhang, P. Luo, and W. Zuo, *Human centric visual analysis with deep learning*. Springer, 2020.

[82] T. Lee and S. Soatto, "Video-based descriptors for object recognition," *Image and Vision Computing*, vol. 29, no. 10, pp. 639–652, 2011.

[83] G. Chunxian, Z. Zhe, and L. Hui, "Hybrid video stabilization for mobile vehicle detection on SURF in aerial surveillance," *Discrete Dynamics in Nature and Society*, vol. 2015, pp. 1–12, 2015.

[84] F. Chen, D. Delannay, and C. Vleeschouwer, "An autonomous framework to produce and distribute personalized team-sport video summaries: A basketball case study," *IEEE Transactions on multimedia*, vol. 13, no. 6, pp. 1381–1394, 2011.

[85] E. Kermani and D. Asemani, "A robust adaptive algorithm of moving object detection for video surveillance," *Eurasip Journal on Image and video processing*, vol. 2014, no. 1, pp. 1–9, 2014.

[86] H. Mansour, P. Nasiopoulos, and V. Krishnamurthy, "Rate and distortion modeling of CGS coded scalable video content," *IEEE Transactions on Multimedia*, vol. 13, no. 2, pp. 165–180, 2010.

[87] A. Halimi, Y. Altmann, N. Dobigeon, and J. Y. Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4153–4162, 2011.

[88] J. Yang, Y. Chen, W. Trappe, and J. Cheng, "Detection and localization of multiple spoofing attackers in wireless networks," *IEEE Transactions on Parallel and Distributed systems*, vol. 24, no. 1, pp. 44–58, 2012.

[89] B. McFee, C. Galleguillos, and G. Lanckriet, “Contextual object localization with multiple kernel nearest neighbor,” *IEEE Transactions on Image Processing*, vol. 20, no. 2, pp. 570–585, 2010.

[90] R. Kwitt, N. Vasconcelos, S. Razzaque, and S. Aylward, “Localizing target structures in ultrasound video–a phantom study,” *Medical image analysis*, vol. 17, no. 7, pp. 712–722, 2013.

[91] T. Shanableh, “Data hiding in MPEG video files using multivariate regression and flexible macroblock ordering,” *IEEE transactions on information forensics and security*, vol. 7, no. 2, pp. 455–464, 2011.

[92] A. Nakhmani and A. Tannenbaum, “Self-crossing detection and location for parametric active contours,” *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3150–3156, 2012.

[93] C. H. Ling, C. W. Lin, C. T. Hsu, and H. Y. M. Liao, “Examplar-based object posture super-resolution using manifold learning,” in *2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2012, pp. 141–145.

[94] M. A. Nicolaou, V. Pavlovic, and M. Pantic, “Dynamic probabilistic CCA for analysis of affective behavior and fusion of continuous annotations,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 7, pp. 1299–1311, 2014.

[95] S. Thudumu, P. Branch, J. Jin, and J. J. Singh, “A comprehensive survey of anomaly detection techniques for high dimensional big data,” *Journal of Big Data*, vol. 7, no. 1, pp. 1–30, 2020.

[96] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection: A survey,” *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.

[97] H. Yang, F. Xie, and Y. Lu, “Clustering and classification based anomaly detection,” in *International conference on fuzzy systems and knowledge discovery*. Springer, 2006, pp. 1082–1091.

[98] M. J. Saberian and N. Vasconcelos, “Multiclass boosting: Margins, codewords, losses, and algorithms.” *Journal of Machine Learning Research*, vol. 20, no. 137, pp. 1–68, 2019.

[99] A. O. Ercan, A. E. Gamal, and L. J. Guibas, “Object tracking in the presence of occlusions using multiple cameras: A sensor network approach,” *ACM Transactions on Sensor Networks (TOSN)*, vol. 9, no. 2, pp. 1–36, 2013.

[100] Y. F. Pan, X. Hou, and C. L. Liu, "A hybrid approach to detect and localize texts in natural scene images," *IEEE transactions on image processing*, vol. 20, no. 3, pp. 800–813, 2010.

[101] X. Yang, H. Liu, and L. J. Latecki, "Contour-based object detection as dominant set computation," *Pattern Recognition*, vol. 45, no. 5, pp. 1927–1936, 2012.

[102] C. Lemaitre, M. Perdoch, A. Rahmoune, J. Matas, and J. Miteran, "Detection and matching of curvilinear structures," *Pattern recognition*, vol. 44, no. 7, pp. 1514–1527, 2011.

[103] A. Walha, A. Wali, and A. M. Alimi, "Video stabilization with moving object detecting and tracking for aerial video surveillance," *Multimedia Tools and Applications*, vol. 74, no. 17, pp. 6745–6767, 2015.

[104] B. Kang, W.-P. Zhu, and J. Yan, "Object detection oriented video reconstruction using compressed sensing," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, pp. 1–15, 2015.

[105] X. Wang, C. Ning, and L. Xu, "Spatiotemporal difference-of-gaussians filters for robust infrared small target tracking in various complex scenes," *Applied Optics*, vol. 54, no. 7, pp. 1573–1586, 2015.

[106] W. C. Hu, C. H. Chen, C. M. Chen, and T. Y. Chen, "Effective moving object detection from videos captured by a moving camera," in *Intelligent Data analysis and its Applications, Volume I.* Springer, 2014, pp. 343–353.

[107] C. H. Chuang, S. C. Cheng, C. C. Chang, and Y. P. P. Chen, "Model-based approach to spatial–temporal sampling of video clips for video object detection by classification," *Journal of Visual Communication and Image Representation*, vol. 25, no. 5, pp. 1018–1030, 2014.

[108] J. Ferryman, D. Hogg, J. Sochman, A. Behera, J. A. Rodriguez-Serrano, S. Worgan, L. Li, V. Leung, M. Evans, and P. Cornic, "Robust abandoned object detection integrating wide area visual surveillance and social context," *Pattern Recognition Letters*, vol. 34, no. 7, pp. 789–798, 2013.

[109] X. Zhao, K. H. Lin, Y. Fu, Y. Hu, Y. Liu, and T. S. Huang, "Text from corners: a novel approach to detect text and caption in videos," *IEEE Transactions on Image Processing*, vol. 20, no. 3, pp. 790–799, 2010.

[110] D. J. Ballantyne, H. Y. Chan, and M. A. Kouritzin, "Novel branching particle method for tracking," in *Signal and Data Processing of Small Targets 2000*, vol. 4048. SPIE, 2000, pp. 277–287.

[111] M. A. Kouritzin, “Residual and stratified branching particle filters,” *Computational Statistics & Data Analysis*, vol. 111, pp. 145–165, 2017.

[112] R. Priyadarshi, A. Thakur, and A. D. Singh, “Performance evaluation space-time interest points using branching particle filters,” in *Proceedings of the Third International Conference on Microelectronics, Computing and Communication Systems*. Springer, 2019, pp. 83–90.

[113] H. A. Abdelali, F. Essannouni, L. Essannouni, and D. Aboutajdine, “A new moving object tracking method using particle filter and probability product kernel,” in *2015 Intelligent Systems and Computer Vision (ISCV)*. IEEE, 2015, pp. 1–6.

[114] E. Bauermeister and D. O’Hagan, “Evaluating the sequential importance re-sampling particle filter for radar target track filtering,” in *International Conference on Radar Systems, 2017*. IET, 2017.

[115] R. Zhou, B. Liu, and J. Teng, “Improved resampling particle filter for maneuvering point target tracking,” *Journal of Applied Sciences*, vol. 13, no. 11, pp. 2045–2050, 2013.

[116] R. Chou, Y. Boers, M. Podt, and M. Geist, “Performance evaluation for particle filters,” in *14th International Conference on Information Fusion*. IEEE, 2011, pp. 1–7.

[117] N. Kumar and D. D. Raj, “Video processing and its applications: A survey,” *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), ISSN*, pp. 2278–6856, 2017.

[118] A. C. Kokaram, “On missing data treatment for degraded video and film archives: a survey and a new bayesian approach,” *IEEE Transactions on Image Processing*, vol. 13, no. 3, pp. 397–415, 2004.

[119] X. Li, J. Han, and S. Kim, “Motion-alert: automatic anomaly detection in massive moving objects,” in *International Conference on Intelligence and Security Informatics*. Springer, 2006, pp. 166–177.

[120] V. Tsakanikas and T. Dagiuklas, “Video surveillance systems-current status and future trends,” *Computers & Electrical Engineering*, vol. 70, pp. 736–753, 2018.

[121] F. Farahi and H. S. Yazdi, “Probabilistic kalman filter for moving object tracking,” *Signal Processing: Image Communication*, vol. 82, pp. 1–16, 2020.

[122] J. Qiu, L. Wang, Y. H. Hu, and Y. Wang, “Two motion models for improving video object tracking performance,” *Computer Vision and Image Understanding*, vol. 195, pp. 1–9, 2020.

[123] S. Jha, C. Seo, E. Yang, and G. P. Joshi, "Real time object detection and tracking system for video surveillance system," *Multimedia Tools and Applications*, vol. 80, no. 3, pp. 3981–3996, 2021.

[124] Y. Shu, C. Xiong, and S. Fan, "Interactive design of intelligent machine vision based on human–computer interaction mode," *Microprocessors and microsystems*, vol. 75, pp. 1–9, 2020.

[125] N. Alqudah and Q. Yaseen, "Machine learning for traffic analysis: a review," *Procedia Computer Science*, vol. 170, pp. 911–916, 2020.

[126] X. Wu, D. Sahoo, and S. C. Hoi, "Recent advances in deep learning for object detection," *Neurocomputing*, vol. 396, pp. 39–64, 2020.

[127] C. L. Chowdhary and D. P. Acharjya, "Segmentation and feature extraction in medical imaging: a systematic review," *Procedia Computer Science*, vol. 167, pp. 26–36, 2020.

[128] S. Kumar and J. S. Yadav, "Video object extraction and its tracking using background subtraction in complex environments," *Perspectives in Science*, vol. 8, pp. 317–322, 2016.

[129] N. Singla, "Motion detection based on frame difference method," *International Journal of Information & Computation Technology*, vol. 4, no. 15, pp. 1559–1565, 2014.

[130] S. Jeeva and M. Sivabalakrishnan, "Survey on background modeling and foreground detection for real time video surveillance," *Procedia Computer Science*, vol. 50, pp. 566–571, 2015.

[131] M. Li, H. Wang, L. Yang, Y. Liang, Z. Shang, and H. Wan, "Fast hybrid dimensionality reduction method for classification based on feature selection and grouped feature extraction," *Expert Systems with Applications*, vol. 150, pp. 1–29, 2020.

[132] A. Mustafa, H. Kim, and A. Hilton, "MSFD: Multi-scale segmentation-based feature detection for wide-baseline scene reconstruction," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1118–1132, 2018.

[133] E. Poongothai and A. Suruliandi, "Global and local oriented gabor texture histogram for person re-identification," *Brazilian Archives of Biology and Technology*, vol. 62, 2019.

[134] J. Wang and W. Zhang, "A survey of corner detection methods," in *2018 2nd International Conference on Electrical Engineering and Automation (ICEEA 2018)*. Atlantis Press, 2018, pp. 214–219.

[135] P. Ram and S. Padmavathi, "Analysis of harris corner detection for color images," in *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*. IEEE, 2016, pp. 405–410.

[136] R. Criado, M. Romance, and A. Sanchez, "Interest point detection in images using complex network analysis," *Journal of Computational and Applied Mathematics*, vol. 236, no. 12, pp. 2975–2980, 2012.

[137] D. Fortun, P. Bouthemy, and C. Kervrann, "Optical flow modeling and computation: A survey," *Computer Vision and Image Understanding*, vol. 134, pp. 1–21, 2015.

[138] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 886–893.

[139] Y. Kim and H. Bang, "Introduction to kalman filter and its applications," *Introduction and Implementations of the Kalman Filter*, vol. 1, pp. 1–16, 2018.

[140] E. Komagal, B. Yogameena, J. Keerthiga, R. Manimozhi, J. Ragavi, G. Nandhini, and G. Surya, "Moving object detection and tracking in wide area surveillance using hungarian-kalman algorithm," *International Journal of Engineering Research & Technology*, vol. 5, no. 9, pp. 1–6, 2017.

[141] K. Saho, "Kalman filter for moving object tracking: Performance analysis and filter design," *Kalman Filters-Theory for Advanced Applications*, pp. 233–252, 2017.

[142] E. Arıcan and T. Aydın, "Object detection with RGB-Ddata using depth oriented gradients," in *Book of Proceedings-International Conference on Engineering and Natural Sciences*, 2017.

[143] X. Qiang, Y. Zhu, and R. Xue, "An improved particle filter for a nonlinear/non-gaussian environment," *IEEE Access*, vol. 7, pp. 151 638–151 651, 2019.

[144] M. Jaward, L. Mihaylova, N. Canagarajah, and D. Bull, "Multiple object tracking using particle filters," in *2006 IEEE Aerospace Conference*. IEEE, 2006, pp. 1–8.

[145] S. K. Weng, C. M. Kuo, and S. K. Tu, "Video object tracking using adaptive kalman filter," *Journal of Visual Communication and Image Representation*, vol. 17, no. 6, pp. 1190–1208, 2006.

[146] A. Kulkarni and M. Vargantwar, "Video based tracking with mean-shift and kalman filter," *International Journal of Engineering Research and Technology*, vol. 3, no. 5, pp. 1–3, 2014.

[147] X. Chen, X. Wang, and J. Xuan, "Tracking multiple moving objects using unscented kalman filtering techniques," *International Conference on Engineering and Applied Science*, 2012.

[148] C. R. Mihalache and B. Apostol, "Hand pose estimation using HOG features from RGB-D data," in *17th International Conference on System Theory, Control and Computing (ICSTCC)*. IEEE, 2013, pp. 356–361.

[149] J. Wang and W. Zhang, "A survey of corner detection methods," in *2018 2nd International Conference on Electrical Engineering and Automation (ICEEA 2018)*. Atlantis Press, 2018, pp. 214–219.

[150] C. Gao, H. Zhu, and Y. Guo, "Analysis and improvement of SUSAN algorithm," *Signal Processing*, vol. 92, no. 10, pp. 2552–2559, 2012.

[151] S. M. M. Kahaki, M. J. Nordin, and A. H. Ashtari, "Contour-based corner detection and classification by using mean projection transform," *Sensors*, vol. 14, no. 3, pp. 4126–4143, 2014.

[152] T. Nayak and N. Bhoi, "Object detection and tracking using watershed segmentation and KLT tracker," *Global Journal of Computer Science and Technology*, vol. 20, no. 1, pp. 1–9, 2020.

[153] G. Majumder, M. K. Bhowmik, and D. Bhatacharjee, "Automatic eye detection using fast corner detector of North East Indian (NEI) face images," *Procedia Technology*, vol. 10, pp. 646–653, 2013.

[154] J. Dou, Q. Qin, and Z. Tu, "Robust image matching based on the information of SIFT," *Optik*, vol. 171, pp. 850–861, 2018.

[155] A. Ignat and I. Pavaloi, "Experiments on iris recognition using SURF descriptors, texture and a repetitive method," *Procedia Computer Science*, vol. 176, pp. 175–184, 2020.

[156] A. Agrahari and R. Ghosh, "Multi-oriented text detection in natural scene images based on the intersection of MSER with the locally binarized image," *Procedia Computer Science*, vol. 171, pp. 322–330, 2020.

[157] H. Sharif and M. Holzel, "A comparison of prefilters in ORB-based object detection," *Pattern Recognition Letters*, vol. 93, pp. 154–161, 2017.

[158] R. Kapoor, R. Gupta, S. Jha, and R. Kumar, "Detection of power quality event using histogram of oriented gradients and support vector machine," *Measurement*, vol. 120, pp. 52–75, 2018.

[159] K. Chu and G. H. Liu, "Image retrieval based on a multi-integration features model," *Mathematical Problems in Engineering*, vol. 2020, no. 3, pp. 1–11, 2020.

[160] K. Acharya and D. Ghoshal, "Detection of a shadow of animated video frames in RGB color space," *Procedia computer science*, vol. 132, pp. 103–108, 2018.

[161] A. Shakarami and H. Tarrah, "An efficient image descriptor for image classification and CBIR," *Optik*, vol. 214, pp. 1–11, 2020.

[162] G. Ciaparrone, F. L. Sanchez, S. Tabik, L. Troiano, R. Tagliaferri, and F. Herrera, "Deep learning in video multi-object tracking: A survey," *Neurocomputing*, vol. 381, pp. 61–88, 2020.

[163] S. Hua and D. C. Anastasiu, "Effective vehicle tracking algorithm for smart traffic networks," in *2019 IEEE International Conference on Service-Oriented System Engineering (SOSE)*. IEEE, 2019, pp. 67–6709.

[164] T. Zhan, K. Yin, J. Xiong, Z. He, and S.-T. Wu, "Augmented reality and virtual reality displays: perspectives and challenges," *Iscience*, vol. 23, no. 8, pp. 1–13, 2020.

[165] M. Tian, Y. Bo, Z. Chen, P. Wu, and C. Yue, "Multi-target tracking method based on improved firefly algorithm optimized particle filter," *Neurocomputing*, vol. 359, pp. 438–448, 2019.

[166] X. Sun, W. Wang, D. Li, B. Zou, and H. Yao, "Object contour tracking via adaptive data-driven kernel," *EURASIP Journal on Advances in Signal Processing*, vol. 2020, no. 1, pp. 1–13, 2020.

[167] S. Kanagamalliga and S. Vasuki, "Contour-based object tracking in video scenes through optical flow and gabor features," *Optik*, vol. 157, pp. 787–797, 2018.

[168] T. B. Schon, F. Lindsten, J. Dahlin, J. Wågberg, C. A. Naesseth, A. Svensson, and L. Dai, "Sequential monte carlo methods for system identification," *IFAC-PapersOnLine*, vol. 48, no. 28, pp. 775–786, 2015.

[169] M. Ahwiadi and W. Wang, "An adaptive particle filter technique for system state estimation and prognosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 6756–6765, 2020.

[170] H. Liu, Y. Liu, and H. Du, "A robust framework for vehicle detection and tracking based on particle filter," in *2015 Joint International Mechanical, Electronic and Information Technology Conference (JIMET-15)*. Atlantis Press, 2015, pp. 803–811.

[171] A. Kumar, G. S. Walia, and K. Sharma, "Real-time visual tracking via multi-cue based adaptive particle filter framework," *Multimedia Tools and Applications*, vol. 79, no. 29, pp. 20 639–20 663, 2020.

[172] D. Yuan, X. Li, Z. He, Q. Liu, and S. Lu, "Visual object tracking with adaptive structural convolutional network," *Knowledge-Based Systems*, vol. 194, pp. 1–24, 2020.

[173] S. M. Hammami and M. Alhammami, "Vision-based system model for detecting violence against children," *MethodsX*, vol. 7, pp. 1–7, 2020.

[174] Z. Li, S. Gao, and K. Nai, "Robust object tracking based on adaptive templates matching via the fusion of multiple features," *Journal of Visual Communication and Image Representation*, vol. 44, pp. 1–20, 2017.

[175] Z. NaNa and Z. Jin, "Optimization of face tracking based on KCF and camshift," *Procedia computer science*, vol. 131, pp. 158–166, 2018.

[176] Z. X. Liu, Y. N. Zou, W. X. Xie, and L. Q. Li, "Multi-target bayes filter with the target detection," *Signal Processing*, vol. 140, pp. 69–76, 2017.

[177] D. Fleet and Y. Weiss, "Optical flow estimation," in *Handbook of mathematical models in computer vision*. Springer, 2006, pp. 237–257.

[178] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *International journal of computer vision*, vol. 12, no. 1, pp. 43–77, 1994.

[179] Y. Liu, H. Yu, C. Gong, and Y. Chen, "A real time expert system for anomaly detection of aerators based on computer vision and surveillance cameras," *Journal of Visual Communication and Image Representation*, vol. 68, pp. 1–29, 2020.

[180] A. Grigorev, O. Severiukhina, and I. Derevitskii, "Anomaly detection using adaptive suppression," *Procedia Computer Science*, vol. 156, pp. 274–282, 2019.

LIST OF PUBLICATIONS

List of Papers Published in Journals

Sr. No.	Title of Published Paper	Journal	No:	Volume Issue	Year	Pages
1	Anomaly Detection in Video Frames : Hybrid Gain Optimized Kalman Filter	Multimedia Tools and Applications, Springer	ISSN: 1380-7501	Vol. 82, Issue 8	2023	1-22
2	Performance Analysis of Branching Particle Filter for Moving Object Tracking and Anomaly Detection	International Journal of Future Generation Communication and Networking, SERSC	ISSN: 2233-7857	Vol. 13, Issue 1	2020	927-939
3	Robust Object Tracking Algorithm using Harris Detector based Bayesian Filters	International Journal of Advanced Science and Technology, SERSC	ISSN: 2005-4238	Vol. 29, Issue 3	2020	3578 - 3590
4	Human Motion Recognition Using Optical Flow Based Particle Filtering	Journal of Image Processing & Pattern Recognition Progress, STM Journals	eISSN: 2394-1995	Vol. 5, Issue 1	2018	43-49
5	Video Anomaly Detection using Kalman Based Support Vector Technique	Current Trends in Signal Processing, STM Journals	ISSN: 2277-6176	Volume 8, Issue 2	2018	4-11
6	Object Tracking Using Finite Element Method and Branching Filter	Current Trends in Signal Processing, STM Journals	ISSN: 2277-6176 (Online), ISSN: 2321-4252 (Print)	Vol. 7, Issue 3	2017	35–40

7	Combined Motion (Direction) and Appearance (Color) Anomaly Detection in Videos	International Journal for Scientific Research & Development	ISSN: 2321-0613	Volume: 5, Issue: 6	2017	831-835
8	Various Motion Cues Detection in Videos	International Journal of Innovative Research in Science, Engineering and Technology	ISSN (Online): 2319-8753 ISSN (Print): 2347-6710	Vol. 6, Issue 8	2017	16572-16578

List of Papers Published in Conferences

Sr. No.	Title of Published Paper	Conference	Year
1	Noise Reduction in Images Using Different Types of Filters	International Conference on Software Technology and Engineering Modules, JCBUST	2017
2	Estimation of Optical Flow Field for Motion Detection in Videos	International Conference on Sustainable Development through Research in Engineering and Management, JCBUST	2016
3	Video Behavior Profiling for Anomaly Detection in Videos	International Conference on Paradigm Shift in Management and Technology, JCBUST, Faridabad	2015
4	Refined Anomaly Detection in Videos	National Conference on New Horizons in Technology for Sustainable Energy and Environment, JCBUST	2017
5	A Survey of Block Matching Algorithms	National Conference on Advances in Mathematics & Computing, JCBUST	2017
6	N-Dimensional Trajectory Clustering	National Conference on Role of Science and Technology towards ‘Make in India’, JCBUST	2016

BRIEF PROFILE OF RESEARCH SCHOLAR

Neetu Gupta, Assistant Professor, Electronics Engineering Department, J.C. Bose University of Science and Technology, YMCA, Faridabad, has an experience of more than 15 years in teaching. She is currently pursuing her Ph.D degree from the same University. Her doctoral research investigates the use of Bayes Filters for anomaly detection in videos. She takes a multidisciplinary approach that encompasses the fields of artificial intelligence, machine learning, computer vision and pattern recognition. During her course of teaching and research work she has *authored and coauthored over 20 research papers in different academic journals and national and international conferences*. She holds a master's degree in Communication Systems and a bachelor's degree in Electronics and Communication Engineering from *Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, Madhya Pradesh*.