

Detecting moving objects in video frames

Makvana Vipul



Department of Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela – 769 008, India

Detecting moving objects in video frames

Dissertation submitted in

May 2014

to the department of

Computer Science and Engineering

of

National Institute of Technology Rourkela

in partial fulfillment of the requirements

for the degree of

Master of Technology

by

Makvana Vipul

(Roll 212CS1085)

under the supervision of

Dr. Pankaj Kumar Sa



Department of Computer Science and Engineering

National Institute of Technology Rourkela

Rourkela – 769 008, India



Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela-769 008, India. www.nitrkl.ac.in

Dr. Pankaj Kumar Sa
Asst. Professor

Certificate

This is to certify that the work in the thesis entitled *Detecting Moving objects in video frames* by *Makvana Vipul*, bearing roll number 212CS1085, is a record of his research work carried out by him under my supervision and guidance in partial fulfillment of the requirements for the award of the degree of *Master of Technology* in *Computer Science and Engineering*.

Pankaj Kumar Sa

Acknowledgement

I wish to express my sincere and heartfelt gratitude towards my guide Prof. Pankaj Kumar Sa, Computer Science Engineering Department, for his supervision and inspiration and above all, help in all regards during the entire duration of my project without which, completion of the project was not possible at all. His guidance has been crucial for giving me a deep insight into the project. I am very much indebted to Prof. S.K.Rath, Head of the Department of CSE for the resources and facilities that were made available to me whenever i needed the same. I would also like to thank all the professors of the department of Computer Science and Engineering, National Institute of Technology, Rourkela, for their constant motivation and guidance.

Many thanks to my comrades and fellow research colleagues. It gives me a sense of happiness to be with you all. Special thanks to Anshuman, Mahesh, Ashok, karthik, Priyesh, Manish, Vijay, Rajesh, and Anoop whose support gave a new breath to my research.

Finally, my heartfelt thanks for their unconditional love and support. Words fail me to express my gratitude to my beloved parents who sacrificed their comfort for my betterment.

Makvana Vipul

Declaration

I, **Makvana Vipul** (Roll No. **212cs1085**) understand that plagiarism is defined as any one or the combination of the following

1. Uncredited verbatim copying of individual sentences, paragraphs or illustrations (such as graphs, diagrams, etc.) from any source, published or unpublished, including the internet.

2. Uncredited improper paraphrasing of pages or paragraphs (changing a few words or phrases, or rearranging the original sentence order).

3. Credited verbatim copying of a major portion of a paper (or thesis chapter) without clear delineation of who did or wrote what. (Source:IEEE, the Institute, May 2014)

I have made sure that all the ideas, expressions, graphs, diagrams, etc., that are not a result of my work, are properly credited. Long phrases or sentences that had to be used verbatim from published literature have been clearly identified using quotation marks.

I affirm that no portion of my work can be considered as plagiarism and I take full responsibility if such a complaint occurs. I understand fully well that the guide of the thesis may not be in a position to check for the possibility of such incidences of plagiarism in this body of work.

Date:

Makvana Vipul

Master of Technology

Computer Science and Engineering

NIT Rourkela

Abstract

Object detection and tracking are challenging critical works in many machine vision applications like traffic control, video surveillance, and person tracking. For detecting moving objects in a video, we use object detection algorithms which compares a static background frame at the pixel level with the current frame. In this thesis, we use the combination of kernel density estimation and modified running average method for efficient motion detection. In this method we use single fixed camera with static background for video. Further we use edge detection algorithm together with morphological operation to improve object detection technique and then combine the output of KDE and MRA method. By using adaptive value in modified running average method for detecting the object, output is less affected by motion.

Keywords: Kernel Density Estimation(KDE), Modified Running Average(MRA), Object Detection, Video Surveillance.

Contents

Certificate	ii
Acknowledgment	iii
Declaration	iv
Abstract	v
List of Figures	viii
List of Tables	ix
1	1
1.1 Introduction	1
1.2 Motivation	5
1.3 Objective	6
1.4 Thesis Layout :	7
2 Background	8
2.1 Introduction	8
2.2 Object Representation :	11
2.3 Object Detection :	14
2.4 Literature Review on Object detection Techniques	16
2.5 Related work:	22
2.5.1 Frame Differencing :	23
2.5.2 Moving Average Filtering :	24

2.5.3	Density Approximation :	24
2.5.4	Running Average :	25
2.6	Combination of KDE and RA	26
2.7	Summary	29
2.8	Results	30
2.8.1	Object Detection Comparison Of traffic video	33
2.8.2	Object Detection Comparison Of Person video	34
3	Conclusions and Future Work	35
3.1	Conclusion	35
3.2	Future Work	36
	Bibliography	37

List of Figures

1.1	The system block diagram	4
2.1	A generic framework for smart video processing algorithms	9
2.2	Object Representation [1]	12
2.3	System Detection method	27

List of Tables

2.1	Analysis of Traffic Control Video of Different Methods	31
2.2	Analysis of Person Walking Video of Different Methods	32
2.3	Comparison Results of Object Detection in Traffic video	33
2.4	Comparison Results of Object Detection in Person walking video .	34

Chapter 1

1.1 Introduction

Video surveillance is a huge exploration subject in computer vision which tries to identify, recognize and tracking the way of objects over an arrangement of image frames and it likewise makes an attempt to comprehend and describe object conduct by supplanting the old customary system for observing cameras by human administrators. Object identification and tracking are exceptionally essential and in addition testing undertakings in numerous workstation vision provisions, in the same way as surveillance, vehicle route and robot route. Object location includes discovering objects in the edge of a video succession caught by camera. For the greater part of the tracking strategy it is fundamental that object location procedure is finished all image outlines or at the time of the object first time appears in a video frame. Object tracking is the system for locating an object or different objects about whether taking images from a video camera.

Video surveillance systems have vastly used in security sensitive areas. The video surveillance history made up of systems with three propagations: 1GSS, 2GSS and 3GSS [2]. 1GSS (first generation surveillance systems) were mostly depended on the analog subsystems of image acquisition, processing and transmission. They amplified eye of human in spatial sense by transferring yields of a few cameras checking set of locales to the presentations in focal controlling room. There are such significant limitations such as, need of high transfer speed,

troublesome filing and recovery of occasions because of huge number of video tape necessities and troublesome online event recognition which just relied on upon human administrators with restricted consideration compass. The surveillance system of second generation (2GSS, 1980 to 2000) was combination in a way that the utilization of both simple and advanced sub frameworks to purpose some inconveniences of its priors. The utilization of the early developments of computerized video handling routines which give help to human administrators to separating out such occasions. The majority work throughout 2GSS is centered on realtime object recognition. Third generation surveillance system (3GSS,2000 -) gives end-to-end advanced frameworks. Image obtaining and handling at level of sensor , correspondence through fixed heterogeneous and mobile systems of broadband and picture storage at main servers formal from ease computerized foundation

Unlike, to past generations, 3GSS has such piece of the picture transforming process which is conveyed towards level of sensor by utilization of good cameras which are fit to program and layer obtained simple picture signs and generate algorithms of image analysis like face identification and motion with assistance of their appended advanced processing parts. A definitive objective of 3GSS is to permit video data which is utilized for the online caution generation to help human administrators and for logged off examination effectively. In place to attain this objective, 3GSS will give smart frameworks that can produce continuous cautions characterized on intricate events and handle distributed capacity and substance based recovery of video information. For making the video surveillance frameworks "smart" needs a quick, solid and powerful algorithms for moving object recognition, grouping, following and event analysis. Beginning from 2GSS, the lot of research work has been given for advancement of these advanced algorithms. Detection of Moving object is an essential task for more examination of the video. This extracts the moving object from the fixed background video frames. It is not just makes a center of consideration for upper level complex calculation but also minimize the time for computation. Normally utilized procedures for object detection are statistical models, background subtraction, optical stream

and temporal differencing. Because of element natural conditions, for example, changing light intensity, waving tree branches in the wind and shadows of object is a troublesome and critical issue that requires a taken care of a good visual surveillance framework.

Classification of Object step arranges locating objects in a predefined classes for example, human, vehicle, creature, mess, and so forth. It is important to recognize objects from one another keeping in mind the end goal to track and break down their activities dependably. Right now, there are two real methodologies of moving object classification, that are shape-based techniques and motion-based techniques [3]. Shape-based strategies make utilization of object's 2D data while motion-based strategies use temporal tracked characteristics of the objects for charactering result. Recognizing characteristic sensation for example, smoke and fire may be fused in object classification parts of visual surveillance frameworks. Locating fire and for raising cautions make human administrators take safety measures in shorter period of time which might spare properties, timberlands and creatures from cataclysmic results.

The following are the step in analysis of video tracking, that could be basically characterized as making of transient correspondence around identified objects from one image frame to another. These techniques give transient distinguishing proof of the segmented areas and provide such information of objects in observed area like, trajectory, direction and speed. The result processed by following step is often used to help and improve motion segmentation, classification of object and furthermore larger amount movement analysis. The last venture of smart video surveillance frameworks is to perceive the movement of objects and make abnormal amount semantic portrayals of their activities. It might essentially be acknowledged as a characterization issue of temporary movement indicators of the objects as indicated by pre-labeled reference indicators speaking to ordinary human movements [4]. The result of these algorithms could be utilized for both giving the human administrator with abnormal amount information to support him to take decisions for additional accuracy and furthermore in shorter period of time and for logged off indexing and looking for the available video information

viably. The developments in the advancement of these algorithms might lead to achievements in provisions that utilize visual surveillance.

There are main three steps in the system of video surveillance :

- 1.locating interesting moving objects in a frame
- 2.tracking of located interested objects from each and every frame, and
3. analysis of trajectory path of object to estimate their behavior in next video frames.

Up to now, the video surveillance system was generally used only for large-scale or high secured companies or military. Although, it will be useful to control the increased crime rate, especially in high-tech cities, took better precautions to decrease criminal activities in securing sensitive places, such as, airports, at the borders of the country, secured government offices, etc. Even people also personally want to seek for their personalized security systems to control their houses or to make secure their valuable things. The general overview is:

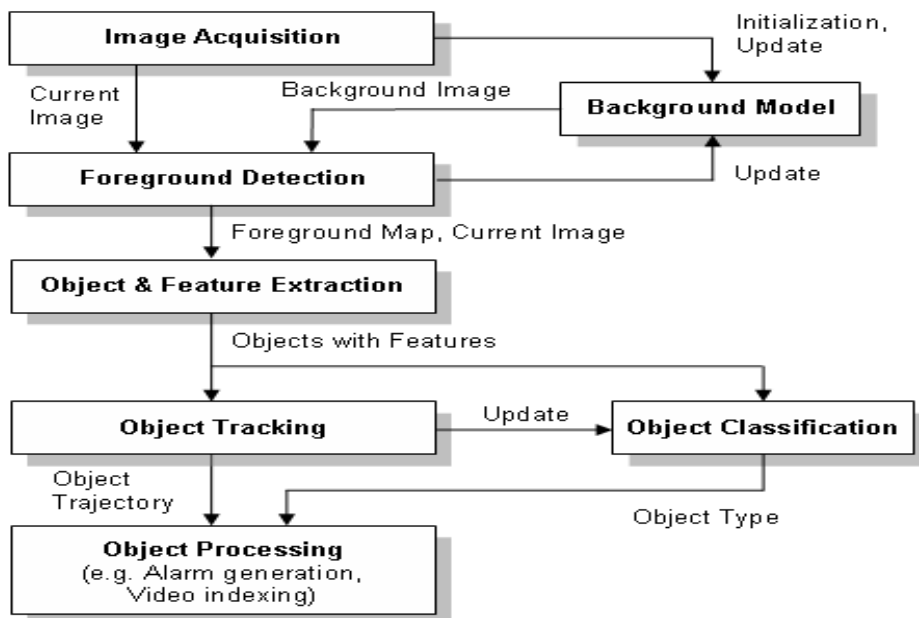


Figure 1.1: The system block diagram

1.2 Motivation

Video surveillance is the best research subject in computer vision for people and vehicles. object tracking is fundamental in numerous computer vision provisions like security and surveillance frameworks, human-machine interfaces, traffic control, feature correspondence/layering. Frequently the fundamental work is to getting the data of moving single or different intrigued questions in grouping of feature casings and preparing this data to gauge the trajectory of items. so, the main motivation behind the video object tracking is to tackle the problem of several important applications such as: Security and surveillance system which is used to recognize people and to provide better level of security using such visual information Retail space instrumentation which is utilized to examine the shopping conduct of diverse clients and to improve building and environment plan, Medical treatment which is utilized to upgrade the nature of the consequence of life or active recuperation of patients and handicapped people.

Here the main objective is to create such visual surveillance framework by replacing the age-old custom strategy for observing by human administrators. The motivation in doing is to outline a feature surveillance framework for correctly choosing the movement location of single or object or different objects in a feature casing and after that doing object tracking.

1.3 Objective

The aim of this thesis is to improve the performance of object detection and tracking by detecting the movement of object in the images of continuous video frames. Automatic tracking of objects can be the foundation for many interesting applications. An accurate and efficient tracking capability at the heart of such a system is essential for building higher level vision-based intelligence. From the previous section it is found that there are many problems in detecting of an object and tracking of objects and also recognition for fixed camera network.

The ultimate goal of the work in this thesis is:

- To set up a system for automatic detecting and tracking of moving Objects with different speed in stationary camera video scenes, which may serve as a foundation for higher level reasoning tasks and applications.
- To make significant improvements in commonly used algorithms for detection. Finally, the aim is to show object detection and tracking based on motion from a static camera.

1.4 Thesis Layout :

The thesis is organized as following :

Chapter 2 : This chapter discusses about the background concepts related to this project work. The chapter also discusses about object representation and some object detection techniques. The literature surveys that have been done during the research work have been discussed here. It also provides a detailed survey of the literature related to motion detection of objects in video frames. Discussion about the existing and some new methods for detection and tracking of objects are done. This methodology and its results are also discussed here.

Chapter 3 : This chapter provides the concluding remarks which can be made to the Project. The scopes for further research are outlined at the end.

Chapter 2

Background

2.1 Introduction

Object tracking in video surveillance framework is an exceptionally important work in the territory of computer vision. Object tracking is the procedure of assessing the way or the path of an object in the video frames. The need for automated video analysis system is increasing; Now a days, Because of inexpensive and high quality video cameras are available, the high-powered computers have created a great job of interest in the algorithms of tracking. The main three key steps in the analysis of video process are, [5] :

- Moving objects detection in video frames.
- Track the detected object or objects from one frame to another.
- Study of tracked object paths to estimate their behaviors.

Digital video processing might be characterizes as the preparing of video by a digital computer [6] . In the memory of a computerized workstation, storage of video frames could be seen as heap of frames alongside the time as pivot (t) with spatial data of each one frame being spoken to by the (x, y) measurement. Figure 2.1 depicts a pictorial perspective of the same.

Mathematically every image frame is matrix of order $i \times j$, and the th image frame be defined as a matrix:

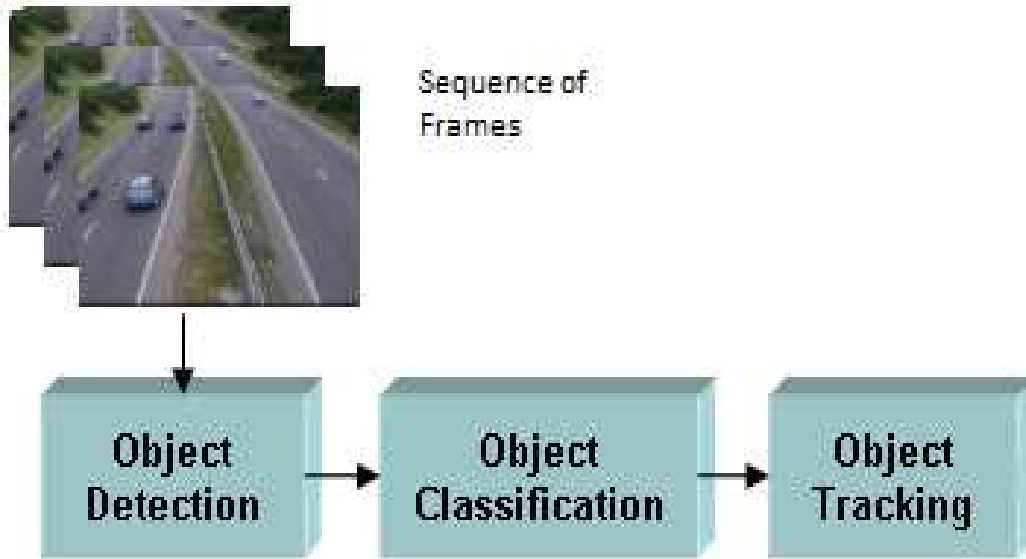


Figure 2.1: A generic framework for smart video processing algorithms

$$f(m, n, t) = \begin{bmatrix} f(0, 0, t) & f(0, 1, t) & \cdots & f(0, j-1, t) \\ f(1, 0, t) & f(1, 1, t) & \cdots & f(1, j-1, t) \\ \vdots & \vdots & \ddots & \vdots \\ f(i-1, 0, t) & f(i-1, 1, t) & \cdots & f(i-1, j-1, t) \end{bmatrix}$$

where j and j is the width and height of the image frame respectively. The pixel intensity or gray value at location (m, n) at time t is denoted by (m, n, t) .

Generally the main use of Track the object applicable in a task of:

- Object recognition based on motion
- Vehicle navigation
- Automated video surveillance
- interaction between Human and computer
- Video indexing
- monitoring of traffic

While tracking, Tracker allocates persistent labels to the interested objects which were being tracked in different images of the video sequence and a tracker gives an object related information, such as area shape or orientation of the object depending on the tracking area. Object Tracking task can become complex due to:

- Because of while converting the 3D images to a 2D images some information loss is occur
- Noise in images
- Full and Partial object occlusions
- object motion is Complex
- Nature of objects is Non rigid or articulated
- Requirements of realtime calculations

We can make the process of tracking simple by providing restriction on the appearance or motion of the objects. Like, most of the algorithms of tracking work with on believing that the movement of object is smooth with no direct changes but it may not occur always. We can further compel the motion of the object as of persistent acceleration or the persistent velocity based on the preliminary information and preliminary knowledge of the object features.

2.2 Object Representation :

In general, Objects can be represented by their features. An object can simply defined as an entity of interest in the video frames for further analysis of image. For example, vehicles on road, fish in an aquarium, boats on the sea, planes, walking people on the road might be useful to track in such a particular kind of domain. So there are various kinds of representations of object shape, which is commonly used for tracking object and after that addresses the joint appearance and shape of object representations which can be described as [1]:

- **Points :** Object can be expressed by a point, which is a centroid (in Figure 2.2(1)) or a set of points (figure 2.2 (2)). The point representation of object is advisable when the object occupied small regions in an image
- **Primitive geometric shape :** Geometric shape means that the shape of te object is expressed by ellipse, rectangle, (in Figure 2.2 (3), (4)). Primitive geometric shape is more relevant for expressing the simple non-rigid objects as well as rigid objects.
- **Object silhouette and contour :** Contour means the boundary of an object(in Figure 2.2 (7), (8)). The inner area of the contour is known as the object silhouette (in Figure 2.2 (9)). These types of representations are relevant for tracking objects having complex non rigid Shapes.
- **Articulated shape models :** The parts of the body which are taken together by joints can be represented as Articulated objects(in Figure 2.2

(5)). Take an example; the person body is composed of articulated object with head, legs, torso, feet, and hands which are connected by the joints.

- **Skeletal model** : Object skeleton be defined as the central axis applied to the object silhouette. This type of model is generally useful as a representation of a shape for identifying interested objects. Skeleton model is used to deal with both rigid and articulated objects. (in figure 2.2 (6))

Object representations. (1) Centroid, (2) multiple points, (3) rectangular patch, (4) Elliptical patch, (5) part based multiple patches, (6) object skeleton, (7) complete Object contour, (8) object contour control points, (9) object silhouette

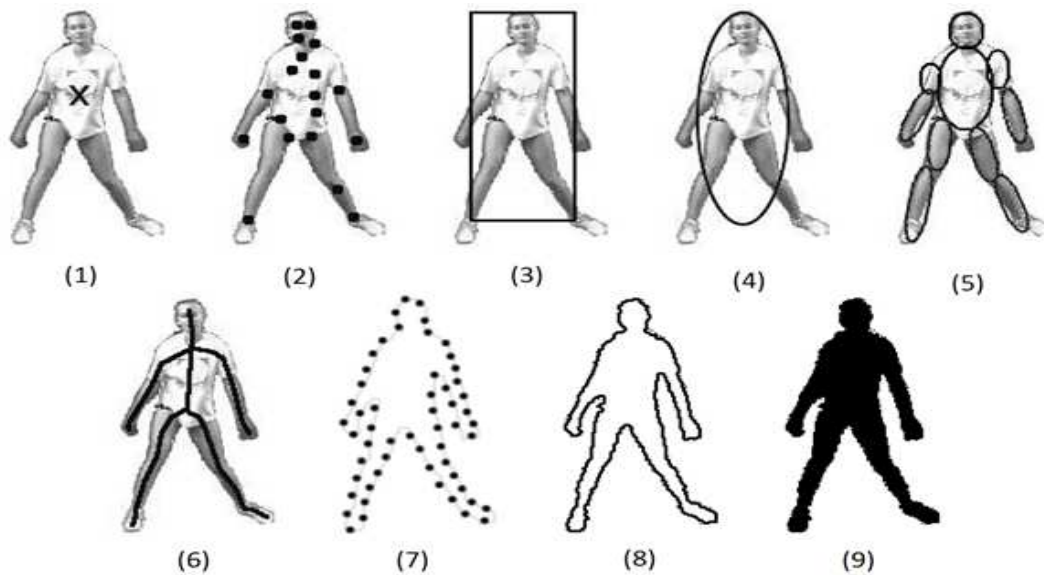


Figure 2.2: Object Representation [1]

There are so many different ways to describe objects based on some appearance features. It should be noted down that shape and appearance feature of object can be blended together for tracking. There are such appearance representation of object which can be useful in tracking can be described as:

- **Probability density estimation of object appearance :** Probability density function can determine the location of an object based on shape feature and use parametric methods like, Gaussian and a mixture of Gaussians, and non-parametric methods like, histograms and Parzen windows. The probability density estimation of an object based on shape features like, texture, color and calculated from the picture regions based on shape models. (a contour or interior ellipse region)
- **Templates :** Templates which can be created using silhouettes or some simple geometric shapes. It provides both appearance and dimensional information of object. Although, template encode object appearance only based on a single view. Thus, this is applicable only for such objects whose location does not change drastically during the tracking process.
- **Active appearance models :** The appearance models can be created generally based on shape and appearance of an object. This can be created by modelling both shape and appearance of an object. Shape of object can be represented by some set of features. Each feature vector is stored by taking texture, color or gradient magnitude feature. These models have learned about shape and appearance of object from such a set of data samples during the training phase.
- **Multi-view appearance models :** this model gives different approaches to encode the different view of the object. One of the approaches is to generate subspace based on the provided view to describe different views of object. Such type of example of subspace approaches like, PCA (is Principal Component Analysis) , ICA (Independent component Analysis). One of the

limitations of this multi view model is that object appearance in all of views needs a so much computation. So it is very time-consuming.

2.3 Object Detection :

Generally, most of the tracking methods do a object detection task in every image frame or at time when object appears first time in a video. The basic idea for detection of an object is locating an object in to the frame only when it appears first time in video. Although, using some temporal information calculated from series of video images, such detection methods are used to reduce the false detection rate. Some object detection methods :

Point detectors :

point detectors generally used to locate interesting points of detecting objects in video frames. The desirable quality of the interested points of object is achieved by illumination changes and changing the camera viewpoint. Generally used point detector methods are: SIFT detector, Harris detector, Moravecs detector, KLT detector

- In Moravecs detector recognize object by locating interested points and using small window for detection. Shifting a window in any direction can give a large change into the intensity value. The main problem of this Moravecs detector is that it gives noisy response because of binary window function and only 45 degree shift is considered. it only select minimum E among edges. So such type of problems can be solved by Harris corner detection.
- SIFT is Scale-invariant feature transform. SIFT detector algorithm was published by David Lowe in 1999. This is a computer vision algorithm to locate and describing local features which are used to detect object. The SIFT features are generally are not affected to

change in illumination, scale and affine distortions. Taking sufficient number of features of images help in robust recognition of object in the cluttered images or partial occlusion. The resulting performance of the object tracking can be further increased by considering extra new SIFT features includes texture , color and edge grouping as well as changing offsets and size of features accordingly. The verification and indexing mechanism can allows all scale and invariant features to be included into the single model. Greatest power might be accomplished by locating numerous diverse characteristic sorts and depending on the indexing and bunching to select those that are most valuable in a specific picture.

- Harris and Stephens uses the differential of corner score by considering the direction directly and gain a better result than Moravec's corner detector by using this instead of shifted patches. The corner score is generally referred to as auto correlation.
- the Kanadelucastomasi (KLT) characteristic tracker is a methodology to characteristic feature extraction. It is proposed essentially with the end goal of managing the issue that customary picture enlistment procedures are by and large excessive. KLT makes utilization of spatial power data to guide the quest for the position that provides the best match. It is speedier than customary systems for inspecting less potential matches between the frames.

Background Subtraction :

Detection of an object could be accomplished by creating to render the background called the background model and afterward discovering intrigued article from model of each one approaching edge. Some change in a picture locale from a background model is reflected in the feature outline. Pixels which are parts of intrigued areas going to change are checked for further handling. The discovery of movements could be accomplished by creating representation of a background foundation and contrasting every new casing

and this representation. This procedure is known as the background subtraction. There are sure techniques for foundation subtraction as examined in the review [1] are Frame differencing Region-based (or) spatial data, Eigen space decomposition and Hidden Markov models (HMM).

Segmentation :

The fundamental objective of the image segmentation calculations is to partition a picture into comparative areas. Each division calculation normally addresses two issues, to decide criteria based on that segmentation of images is doing and the technique for attaining effective dividing. Different division methods those are pertinent to question following are: image segmentation using Graph-Cuts (Normalized cuts), mean shift clustering and Active contours.

Supervised classifiers :

Object identification is possible by taking in diverse views of an object naturally from the set of samples by method for directed taking in component. Taking in of diverse object perspectives permits the necessity of putting away a complete set of layouts. From a set of given taking in learning samples and from the set of inputs supervised learning methods generate a desire outputs.

2.4 Literature Review on Object detection Techniques

- Elgammal, Ahmed, et al. [3] proposed a non-parametric based kernel density estimation method for developing statistical representation of foreground and scene background areas in video surveillance task. the probability density function related to the fore ground or background need not follow

the parametric form which are known. The background model is taking a recent intensity value of pixel and estimating the pdf of intensity value of pixel directly. This model obtains a detection of moving object cluttered background and monitor a situation of scene background is not totally stationary but small motions are there such as, moving leaves or branches of tree. We likewise utilized kernel estimation procedures for displaying the foreground areas. We indicated that this system is general methodology for demonstrating homogeneous color districts. We presented the representation of individuals that spatially restricts color properties in such a manner that relates to their garments. In light of this type of representation, we exhibited general probabilistic system which uses maximum-likelihood estimation to gauge the best game plan for individuals in gathering so as to fragment the foreground locales comparing to this gathering.

The method which create and control the model of object occlusion that can be utilized into the same segmentation methodology. This paper mainly focuses on solving the problem of a statistical expression of background that provides a sensitive moving object detection into the images, but very less effective to natural scene variations we likewise use general nonparametric kernel density methods for building these measurable representations of foreground and background. These methods assess the pdf straightforwardly from the information without any presumptions of the basic appropriations.

- Singh, Abhishek, et al. [7] provide a method containing background subtraction using GMM (Gaussian Mixture Mode) in low contrast frames taken by fixed camera. By using K-means algorithm and Expectation maximization method this technique update Mixture model parameters. This technique gives better result as comparison with already existing methods only either EM or K-means methods in such situations also where images having lower contrast. Because of the intrinsic advantage of utilizing a Gaussian Mixture Model, our framework can manage multi-modal distributions and adjust to lighting variations. The framework

has high potential to be utilized within provisions including military cover, recognition and tracking of balls and so forth in game occasions, tracking of hazed out objects at an extensive separation, around others. To further enhance the execution of the tracker, we are concentrating on two ranges - speed and accuracy. We are utilizing an estimate of the EM calculation to spare on calculation time at the degraded accuracy. For improving the accuracy of the recognized object, implement the split and merge method to identify the Gaussians in a Mixture model more accurately. Additionally, a self-learning method to adaptively focus the ideal number of Gaussians might make the framework more versatile.

- Karaulova, I. A, et al. [8] give a new approach of hierarchical model for tracking of people in sample video frames based on human dynamics. The model is prepared utilizing true information from a gathering of individuals. Kinematics is encoded by utilizing Hierarchical PCA (Principal Component Analysis) , and progress are encoded using HMM (Hidden Markov Models). The highest point of the hierarchy holds data about the entire body. The hierarchy at the lower levels hold more point by point data about conceivable position of some subpart of the body. At the point of tracking, the hierarchy at the lower levels are indicated to enhance accuracy. In this article we portray our model and present examinations that show we can retrieve 3d skeletons from 2d pictures in a perspective freeway and tracking individuals the framework was not prepared on.
- Makris et al. [9] defines the issue of identify the person highways which are most frequently used from the sequences of video frames of outdoor environment. Specifically, marked paths give a productive means for compressing the trajectory information for the purpose of logging. Moreover, the model could be utilized to figure a probabilistic prediction of the pedestrians area numerous time steps ahead, and to support the recognition of unordinary behaviour distinguished as atypical object

movement. This paper has exhibited the common sense of building spatial models focused around the examination of trajectory information removed from picture sequences. The models have been demonstrated to be significant for financially encoding the course emulated by an object in the picture, minimizing the trajectory information down to a solitary name connected with each one course. However surveillance tracking calculations give a next step based on local prediction to help the correspondence transform in the following picture frames, encoding the course and path information helps forecast over numerous time steps, and may be especially helpful for predicting over a few types of impediment in the scene (e.g. parked vehicle).

The path hub predictions created from the courses and paths are confined by the number of trajectories accessible for taking in and more dependable statistics might require any longer preparing periods (i.e. more trajectories). Indeed, it is likely that we might need to segment the course taking in into distinctive time periods (e.g. every hour), as the detail are not stationary about time. The representation of the path models is focused around groupings on bunch focuses and linear interpolation is performed at whatever point is needed. Despite the fact that the accuracy of the results is tasteful, we think about the utilization of cubic splines rather, that they will give more faultless models. The classification procedure exhibited in the results does arrange trajectories that fall outside the current learned state of the model. However to focus a genuine classification of such an occasion requires a more extended time set of perceptions. We think about increasing the probabilistic model of path use to speak to a Markov process, which encodes the history of tracked object.

- Mohan, et al. [10] propose a simple example based method for recognizing objects in a still images by using components. The framework is organized with four different example-based detectors that are prepared to independently find the four parts of the human body: the head, legs, left arm,

and right arm. After confirmed that these parts are available in the correct geometric arrangement, a second example-based classifier consolidates the outputs of the part detectors to characterize an example as either a person or a nonperson. We can define such hierarchical design, in which study of numerous stages occurs, an Adaptive Combination of Classifiers (ACC). From the results that show that this framework performs fundamentally superior to a comparable full-body individual indicator. This proposes that the change in execution is because of the segment based methodology and the ACC information characterization construction modeling. The calculation is additionally more powerful than the full-body individual discovery technique in that it is fit for recognizing partially occluded perspectives of persons and persons whose body parts have little appear differently in relation to the background.

In this paper, the comparison between ACC (Adaptive Combination of Classifiers), VCC (Voting Combination of Classifiers) and whole body detection methodology. The algorithm was run on the databases of several test images on different threshold values. The outputs of this algorithm are plotted and recorded at ROC (receiver operating characteristic) curves and ROC curve of people recognition system achieve trade-off between false detections and accuracy which is essential for every detector. The detailed study of ROC curves shows that people detection based on component based system gives better performance as compared to other baseline system and the baseline system should be used same SVM classifier and image representation methods which was component detector used in component-based method. Thus the notable improvement achieved using combined component classifiers. So, this component based framework work well in variable intensity of lights and in noise in an image than the full body person detector and also recognizes partial occlusion of people. we have introduced a component-based individual detection framework for static pictures that can identify frontal, back, somewhat rotated and partial occlusion of people in jumbled scenes without accepting any from the earlier

information concerning the images of video frames.

- Most scientists have forsaken non-versatile methods for back grounding in light of the fact that the manual introduction must be required. Without introduction, blunders out of sight will gather over time, making this technique helpful just in profoundly regulated, fleeting following requisitions and without any noteworthy changes in the scene. Lipton et al. [6] proposed to utilize a greatest inter frame difference and describe a method for obtaining interested moving object from a real time video frames, categorize them into predefined image based techniques and after that tracking them. However this leaves "ghosts" in the ensuing frame and if the article experiences noteworthy movement in each one casing then the item was and leaves extensive locales of the article undetected.
- Most of the backgrounding systems working with persistently estimating a measurable model of the variable for every pixel in the frame. A typical system for versatile backgrounding work like, averaging the frame picture outlines over time and making a foundation estimate which is for the most part like the current static scene with the exception of where the article movement happens. While this is the circumstances where object move consistently and the background is unmistakable for a critical allotment of the time, it is not powerful to scenes with numerous moving items, especially in the event that they move gradually. It additionally can't deal with the bimodal backgrounds, recuperates gradually when the backgrounds is not secured, and have a solitary decided ahead of time limit for the whole scene. One intriguing endeavor to meet these challenges is W4 [11], which joined its estimation of the maximum value, minimum value, and the maximum inter frame difference per pixel in a video frame.
- A novel feature of the object tracking methodology based on Markov random field and kernel density estimation [12], which might be used in both programmed and self-loader segmentation. Based on the nonparametric

model background and each object are recognized by the kernel density estimation. Utilizing the maximum likelihood measure, every pixel in the frame is initially arranged into an object or background in a data feature outline. The Markov random model that appropriately demonstrates spatial smoothness is specifically used to refine the arrangement result for more faultless video objects. The non-parametric models are overhauled and engendered throughout the entire following procedure. This methodology indicates that it can productively track feature objects with great visual quality.

- In pictures, the pixel intensity is the most usually utilized characteristics as a part of background demonstrating. On the off chance that we screen the pixel intensity esteem about whether in a totally static background, then the power of a pixel could be demonstrated sensibly with a Gaussian distribution $N(\mu, \sigma^2)$, gave that the picture commotion about whether might be displayed by zero mean Gaussian distribution. The Gaussian distribution model for pixel intensity quality is an underlying model for some background subtraction strategies. For instance, one of the least complex foundation subtraction method is to compute a normal picture of the feature scene, subtract every new arriving frame from this picture, and chooses one limit esteem and focused around that come about will be produced. This essential Gaussian model can change to abate changes into scene (for instance, steady brightening progressions) and recursively upgrading the model by utilizing a basic adaptive filter. This fundamental adaptive model is utilized as a part of [13]; likewise, Kalman filtering for adjustment is utilized with [14]- [15].

2.5 Related work:

The Performance of a mechanized visual surveillance framework considerably relies on upon its capability to locate moving objects in the video arrangements of frames. An ensuing activity, for example, tracking, dissecting the motion or

distinguishing objects or persons, requires an exact extraction of the foreground objects, to make moving object detection a significant task of the framework. The locating issue of progressions in a scene could be portrayed as follows: Images of the same frame are gained in time with a static camera and the objective is to locate changes between successive frames. The Pixels of pictures that have a noteworthy distinction contrasted with the past ones are checked as foreground pixels. With a specific end goal to choose whether a few locales in a frame are foreground or not, there ought to be a various methods to choose it. This strategy ought to likewise have the capacity to catch and store essential background data. Such change, which is brought on by another object, ought to be identified by this method. Some object detection techniques which are described as:

2.5.1 Frame Differencing :

The easiest technique for moving object identification is frame differencing. The model for the background is basically equivalent to the past frame.

$$m(m, n, t) = \begin{cases} 0 ; & \text{if } |I(m, n, t) - I(m, n, t - 1)| < th \\ 1 ; & \text{otherwise} \end{cases}$$

In the above equation, $I(m, n, t)$ is the intensity at pixel area (m, n) at time t , th is the edge quality and $m(m, n, t)$ is the change mask acquired in the wake of thresholding.

Based on, utilizing the past frame, a single frame, which does not incorporate any moving objects, can additionally be utilized as a background frame to recognize moving objects. In spite of the fact that this strategy is by and large quick and has adjustment capability to make the progressions in the scene, it has a moderately less execution in dynamic scene conditions and its comes about are extremely delicate to the limit value th . Moreover, in view of a solitary limit value, this technique can't manage multi-modal conveyances [16].

2.5.2 Moving Average Filtering :

In this method, the background frame as a kind of perspective is built by calculating the mean estimation of the past N frames. A change mask is obtained as follows :

$$m(m, n, t) = \begin{cases} 0; & \text{if } |I(m, n, t) - I_r| < th \\ 1; & \text{otherwise} \end{cases}$$

The background model update equation is ,

$$I_r, t = \alpha * I(m, n, t - 1) + (1 - \alpha) * I_r, t - 1 \quad (2.1)$$

As in the frame differencing strategy, the mask, $m(m, n, t)$ is gained after the thresholding by th . In the redesign comparison, α is the taking in parameter. Moving average separating additionally disintegrate from limit affectability and can't manage multi-modal appropriations.

2.5.3 Density Approximation :

- **Kernel Density Estimation (KDE):**

Density estimation is a basic concept which is used in statistics and widely used in computer vision research. Density estimation can be defined as the construction of an unknown density function from the observed data or samples. One approach of the density estimation which uses parametric methods and ultimate goal is to use probability density model which is known and make it suitable for the data samples. Parametrically estimated density requires a priori knowledge of the density function and relies on the model specification. This task is difficult as prior knowledge is often unknown. The problem of parametric density estimation is the accuracy and flexibility.

On the other side for non-parametric methods are not dependent on any typically parameters. Because of the estimated density is dependent only on the structure of the data, the kernel density estimation can work with any density function. Kernel density model is a non-parametric statistical model which estimates the probability density function (pdf) from the finite set of sample data

and based on this data sample interfaces the general problem of data smoothing.

The underlying probability density function (pdf) is defined as :

$$f(x) = \sum_i \alpha_i K(x - x_i) \quad (2.2)$$

[3] where K is the kernel function which can generally be a, Gaussian function centered at sample data point.

x_i , α_i is the weighting co-efficient of the kernel function. For the given data samples $\{ x_i \}$ $i= 1,2,3, \dots, N$ and

the distribution with density function $p(x)$, the background probability density function can be estimated as,

$$P(x) = \frac{1}{N} * \sum_{i=1}^N K_{\sigma}(x_t - x_i) \quad (2.3)$$

$$\sigma = \frac{m}{0.68 * \sqrt{2}} \quad (2.4)$$

[3] Where K_{σ} is a kernel function with bandwidth σ and m is the median of $|x_i - x_{i-1}|$ for $i = 1, 2, 3, \dots, N - 1$

From the above background probability density function $p(x)$, we can find background and foreground pixels based on some threshold value decided for all video image frames. If $p(x) < th$ then it belongs to foreground else it belongs to background. So, locating objects in to the video frames and then estimate its location in the consecutive frames based on the calculations.

2.5.4 Running Average :

In the running average method, the previous background frame $B_{t-1}(x,y)$ and new incoming frame $I_t(x,y)$ are composed together to achieve current background image. The adaptive background model is attained as

$$B_t(x, y) = (1 - \beta) * B_{t-1}(x, y) + \beta * I_t(x, y) \quad (2.5)$$

[17] where β is an adjustable parameter and its value should be adaptive depending on the motion speed. For larger β can leads to faster background changes and smaller β means that the background changes slowly.

The binary motion detection mask $D(x, y)$ can be calculated based on the background frame and the current frame and then compare this value with the global threshold to decide moving objects in to the frame. The motion detection mask can be defined as:

$$D(x, y) = \begin{cases} 1; & \text{if } |I_t(x, y) - B_t(x, y)| > th \\ 0; & \text{otherwise} \end{cases} \quad [17]$$

Where $I_t(x, y)$ is current video frame and $B_t(x, y)$ is current background model and th is an experimental threshold to decide pixel in foreground or in background.

The RA method is not that much accurate for extracting moving objects because the parameter value of β is a fixed value for all pixels in all the frames. For better result the value of β should be adaptive and changing based on the motion speed of the object. For slow motion regions the background should be updated accordingly slowly and β should be small and for faster moving regions background should be updated faster so, β should be large. To achieve that goal we can use cross-correlation values of two consecutive images and assign this value to β for each frame of video images.

The cross-correlation between two images can be defined as

$$C.C(m, n) = \frac{m.n}{|m|.|n|} \quad (2.6)$$

Where m and n are two consecutive images.

So, if the cross-correlation value between two images is small then background should be updated faster for that frame else for large value of cross-correlation, the background should be updated slowly.

2.6 Combination of KDE and RA

This method is the combination of kernel density estimation model and running average method of video image frames. Here first the object can be detected

using the kernel density estimation method and compare it with the canny edge detection method and do some operation on it to get true and complete edges of the object and improve the result output. However, there are some extra edges in the output frames. So our main task is to remove it from the output to get better results. After that this result is compare with the running average method of object detection and after that do some morphological operation on it to get better extracted object detection result. The combined kernel density and running average methods can be work efficiently. The Detection model is shown in figure.

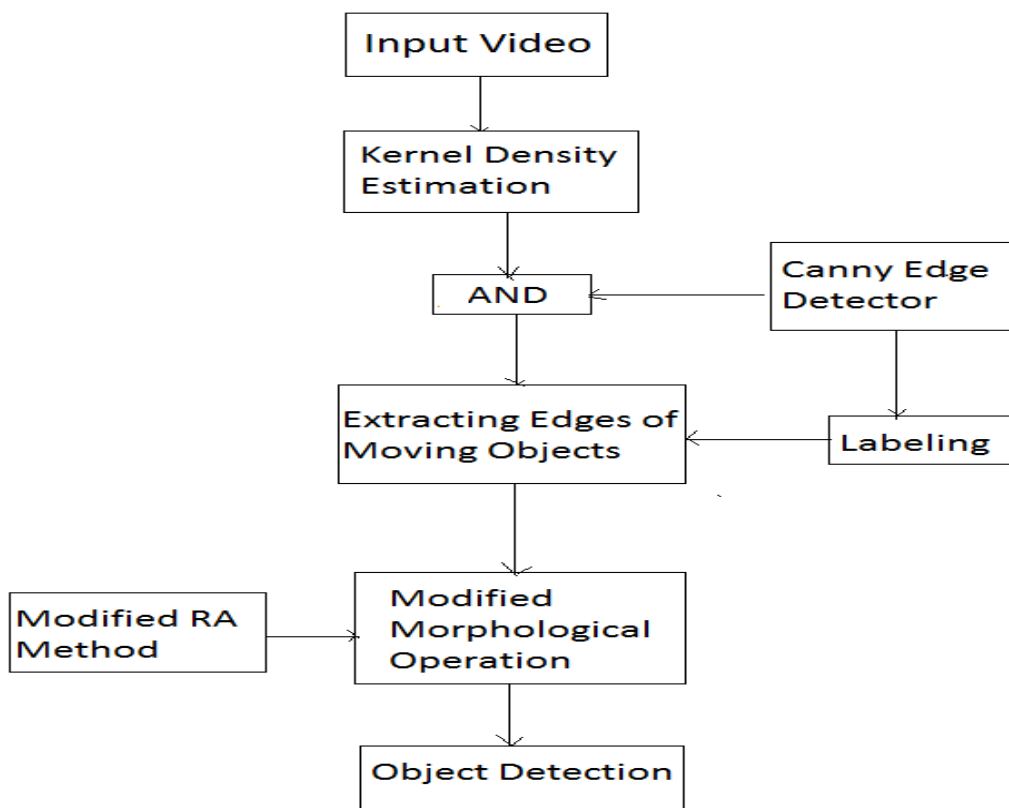


Figure 2.3: System Detection method

The algorithm steps can be as follows:

- (i) The KDE is implemented to incoming frame(F_t) to obtain primary binary motion detection mask(MDM_t).

- (ii) The canny edge detector(C_t) is applied to extract edges of frame F_t . canny edge detector is very robust to noise and used to extract weak edges from the frames.
- (iii) The pixel wise AND operation done between MDM_t and C_t to extract edges of moving objects (ME_t).
- (iv) The edges of the coming frame are labeled as LE_t and combined it with the moving objects edges ME_t . To extract true and complete edges of moving objects finding locations of pixels having the same label in ME_t and its corresponding location labeled in C_t and stored in CLE_t .
- (v) To get more efficient result use Running Average method to F_t and store at RA_t .
- (vi) CLE_t and RA_t is combined and doing some morphological operations to get result of extracted object. For morphological operation calculate a motion ratio which is defined as:

$$\frac{\sum_{c=i}^j MRA_t(r, c)}{j - i + 1} \quad (2.7)$$

Where i and j are the column indices of the two pixels which having the same labels in the same r row of matrix CLE_t . This operation is extracting the whole complete area of the moving object. In each row the pixel between the two same labeled pixels are treated as moving one only if the motion ratio between these two pixels is above a predefined threshold.

In order to do the comparison of different methods, some quantitative metrics are used which are defined as follows:

- (i) **True Positive(tp)**: Correctly Identified Pixels.
- (ii) **False Positive(fp)**: InCorrectly Rejected Pixels.
- (iii) **False Negative(fn)**: InCorrectly Identified Pixels.

So, Similarity, Recall and Precision can be calculated as:

$$Similarity = \frac{tp}{tp + fp + fn} \quad (2.8)$$

$$Recall = \frac{tp}{(tp + fn)} \quad (2.9)$$

$$Precision = \frac{tp}{(tp + fp)} \quad (2.10)$$

2.7 Summary

While doing research work, there are so many challenging problems of detection of moving objects. Because of the speed variation of object it becomes difficult to detect it. This algorithm not only extracts the moving object accurately using a canny edge detector, but also it relatively improves the processing rate. In RA method using the adaptive updating parameter gives a better improved result.

2.8 Results

Two sample videos are used as inputs for this algorithm. They are :-

- **Traffic Sequence** This video of a traffic scene was captured using a fix camera. The cars are moving with the different speed along a straight road. our goal is to detect this oject in a video frames. compare this output images using different methods.
- **Person tracking** This video of a person moving randomly along a scene. Our aim is to detect this person. The output shows the comparison between the different image frame of video.

Table 2.1: Analysis of Traffic Control Video of Different Methods









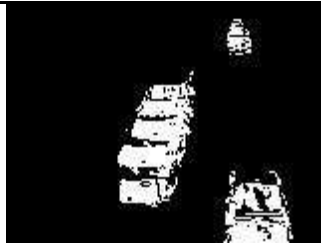



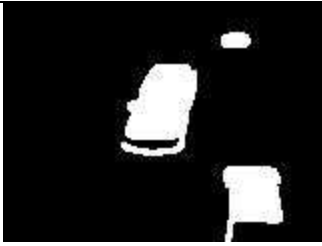

















	Frame no.39	Frame no.40	Frame no.41
Original Frame			
Ground truth image			
Kernel Density Estimation			
Modified Running Average			
Combined KDE and MRA			

Table 2.2: Analysis of Person Walking Video of Different Methods

	Frame no.49	Frame no.50	Frame no.51
Original Frame			
Ground truth image			
Kernel Density Estimation			
Modified Running Average			
Combined KDE and MRA			

2.8.1 Object Detection Comparison Of traffic video

Table 2.3: Comparison Results of Object Detection in Traffic video

Different Detecting Algorithms				
Frame No.	Evaluation	KDE	MRA	Proposed Method
39	Similarity	0.4628	0.1847	0.7715
	Recall	0.5202	0.2205	0.7761
	Precision	0.8074	0.5316	0.9924
40	Similarity	0.4961	0.1768	0.8434
	Recall	0.5788	0.2075	0.8519
	Precision	0.7765	0.5443	0.9883
41	Similarity	0.4114	0.1973	0.8723
	Recall	0.5400	0.2453	0.8727
	Precision	0.6333	0.5017	0.9895

2.8.2 Object Detection Comparison Of Person video

Table 2.4: Comparison Results of Object Detection in Person walking video

Different Detecting Algorithms				
Frame No.	Evaluation	KDE	MRA	Proposed Method
49	Similarity	0.1345	0.1543	0.7178
	Recall	0.1410	0.1667	0.7607
	Precision	0.7442	0.6753	0.9271
50	Similarity	0.1413	0.1636	0.6908
	Recall	0.1439	0.1673	0.7712
	Precision	0.8837	0.8824	0.9089
51	Similarity	0.1561	0.1340	0.6217
	Recall	0.1589	0.1408	0.7384
	Precision	0.7492	0.7953	0.8972

Chapter 3

Conclusions and Future Work

3.1 Conclusion

In recent years, there is significant advancement in object detection methods with the improvement in object features based on different algorithms. Real time deployment of the algorithm desires maximum accuracy with less complexity, which makes the problem still open and needs significant research. This prompted the improvement of object detection calculations. After detecting the object from video frames, the next step is to track that object in next upcoming video frames or make some analysis process of object based on the methodology. Here, the methodology of detecting moving object from a video is described. So, main objective is to efficiently detect the moving object. For this first applying the kernel density estimation method and then compare it with canny edge detector to obtain the edges of moving object. For finding the true complete edges of the moving object with the help of connected components. For making the result more accurate we use running average mechanism with the help of morphological operation of kernel density estimation and by running average method we get more accurate result.

3.2 Future Work

The research work can be extended to many aspects by using other methodologies which can make the object detection process more accurate. In the situation where shadow of object is present, the combined method of KDE & MRA method wont work in some situtation. So we can add other steps to make object detection process more efficient and also we can extend this detection process to color images.

Bibliography

- [1] Alper Yilmaz, Omar Javed, and Mubarak Shah. Object tracking: A survey. *Acm computing surveys (CSUR)*, 38(4):13, 2006.
- [2] Franco Oberti, Giancarlo Ferrari, and Carlo S Regazzoni. A comparison between continuous and burst, recognition driven transmission policies in distributed 3gss. In *Video-Based Surveillance Systems*, pages 267–278. Springer, 2002.
- [3] Ahmed Elgammal, Ramani Duraiswami, David Harwood, and Larry S Davis. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. *Proceedings of the IEEE*, 90(7):1151–1163, 2002.
- [4] Chris Stauffer and W Eric L Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2. IEEE, 1999.
- [5] Rafael C Gonzalez, Richard Eugene Woods, and Steven L Eddins. *Digital image processing using MATLAB*. Pearson Education India, 2004.
- [6] Alan J Lipton, Hironobu Fujiyoshi, and Raju S Patil. Moving target classification and tracking from real-time video. In *Applications of Computer Vision, 1998. WACV'98. Proceedings., Fourth IEEE Workshop on*, pages 8–14. IEEE, 1998.
- [7] Abhishek Singh, Padmini Jaikumar, Suman K Mitra, Manjunath V Joshi, and Asim Banerjee. Detection and tracking of objects in low contrast conditions.

- In *IEEE National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), Gandhinagar, India*, pages 98–103, 2008.
- [8] IA Karaulova, Peter M Hall, and A David Marshall. A hierarchical model of dynamics for tracking people with a single video camera. In *BMVC*, pages 1–10, 2000.
- [9] Dimitrios Makris and Tim Ellis. Path detection in video surveillance. *Image and Vision Computing*, 20(12):895–903, 2002.
- [10] Anuj Mohan, Constantine Papageorgiou, and Tomaso Poggio. Example-based object detection in images by components. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(4):349–361, 2001.
- [11] Ismail Haritaoglu, David Harwood, and Larry S. Davis. W 4: Real-time surveillance of people and their activities. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):809–830, 2000.
- [12] Zhi Liu, Ran Shi, Liquan Shen, Yinzhu Xue, King Ngi Ngan, and Zhaoyang Zhang. Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut. *Multimedia, IEEE Transactions on*, 14(4):1275–1289, 2012.
- [13] Christopher Richard Wren, Ali Azarbayejani, Trevor Darrell, and Alex Paul Pentland. Pfinder: Real-time tracking of the human body. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):780–785, 1997.
- [14] Klaus-Peter Karmann and Achim Von Brandt. Method for detecting and tracking moving objects in a digital image sequence having a stationary background, July 23 1991. US Patent 5,034,986.
- [15] Dieter Koller, J Weber, T Huang, J Malik, G Ogasawara, B Rao, and S Russell. Towards robust automatic traffic scene analysis in real-time. In *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision &*

-
- Image Processing., Proceedings of the 12th IAPR International Conference on*, volume 1, pages 126–131. IEEE, 1994.
- [16] Massimo Piccardi. Background subtraction techniques: a review. In *Systems, man and cybernetics, 2004 IEEE international conference on*, volume 4, pages 3099–3104. IEEE, 2004.
- [17] Shih-Chia Huang. An advanced motion detection algorithm with video quality analysis for video surveillance systems. *Circuits and Systems for Video Technology, IEEE Transactions on*, 21(1):1–14, 2011.
- [18] Ismail Haritaoglu, David Harwood, and Larry S Davis. W 4 s: A real-time system for detecting and tracking people in 2 1/2d. In *Computer Vision ECCV'98*, pages 877–892. Springer, 1998.
- [19] Kalyan Kumar Hati, Pankaj Kumar Sa, and Banshidhar Majhi. Lobs: Local background subtracter for video surveillance. In *Microelectronics and Electronics (PrimeAsia), 2012 Asia Pacific Conference on Postgraduate Research in*, pages 29–34. IEEE, 2012.
- [20] Bijan Shoushtarian and Helmut E Bez. A practical adaptive approach for dynamic background subtraction using an invariant colour model and object tracking. *Pattern Recognition Letters*, 26(1):5–26, 2005.
- [21] Vijay Mahadevan and Nuno Vasconcelos. Background subtraction in highly dynamic scenes. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–6. IEEE, 2008.
- [22] P Wayne Power and Johann A Schoonees. Understanding background mixture models for foreground segmentation. In *Proceedings image and vision computing New Zealand*, volume 2002, pages 10–11, 2002.
- [23] Makito Seki, Toshikazu Wada, Hideto Fujiwara, and Kazuhiko Sumi. Background subtraction based on cooccurrence of image variations. In

-
- Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–65. IEEE, 2003.
- [24] Daniel Gutchess, M Trajkovics, Eric Cohen-Solal, Damian Lyons, and Anil K Jain. A background model initialization algorithm for video surveillance. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 733–740. IEEE, 2001.
- [25] J Cezar Silveira Jacques, Claudio Rosito Jung, and S Raupp Musse. Background subtraction and shadow detection in grayscale video sequences. In *Computer Graphics and Image Processing, 2005. SIBGRAPI 2005. 18th Brazilian Symposium on*, pages 189–196. IEEE, 2005.
- [26] Horng-Horng Lin, Tyng-Luh Liu, and Jen-Hui Chuang. Learning a scene background model via classification. *Signal Processing, IEEE Transactions on*, 57(5):1641–1654, 2009.
- [27] Du-Ming Tsai and Shia-Chih Lai. Independent component analysis-based background subtraction for indoor surveillance. *Image Processing, IEEE Transactions on*, 18(1):158–167, 2009.
- [28] Olivier Barnich and Marc Van Droogenbroeck. Vibe: A universal background subtraction algorithm for video sequences. *Image Processing, IEEE Transactions on*, 20(6):1709–1724, 2011.
- [29] Milan Sonka, Vaclav Hlavac, Roger Boyle, et al. *Image processing, analysis, and machine vision*, volume 3. Thomson Toronto, 2008.
- [30] Liang Wang, Weiming Hu, and Tieniu Tan. Recent developments in human motion analysis. *Pattern recognition*, 36(3):585–601, 2003.
- [31] W Eric L Grimson, Chris Stauffer, Raquel Romano, and Lily Lee. Using adaptive tracking to classify and monitor activities in a site. In *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*, pages 22–29. IEEE, 1998.