A Survey on Object Detection and Tracking Algorithms

Rupesh Kumar Rout

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

A Survey on Object Detection and Tracking Algorithms

Thesis submitted in

June 2013

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 Technologyy

by

Rupesh Kumar Rout

(Roll 211CS1049)

Department of Computer Science and Engineering

National Institute of Technology Rourkela

Rourkela – 769 008, India

Acknowledgment

I am grateful to numerous local and global peers who have contributed towards shaping this thesis. I am very much obliged to Prof. B.Majhi for his guideline, advice and support during my thesis work.I am very much indebted to Prof. Ashok Kumar Turuk, Head-CSE, for his con-tenuous encouragement and support. He is always ready to help with a smile. I am also thankful to all the professors of the department for their support. I am really thankful to my all friends. My sincere thanks to everyone who has provided me with kind words, a welcome ear, new ideas, useful criticism, or their invaluable time, I am truly indebted. I must acknowledge the academic resources that I have got from NIT Rourkela. I would like to thank administrative and technical staff members of the Department who have been kind enough to advise and help in their respective roles. Last, but not the least, I would like to dedicate this thesis to my family, for their love, patience, and understanding.

Rupesh Kumar Rout

Abstract

Object detection and tracking are important and challenging tasks in many computer vision applications such as surveillance, vehicle navigation, and autonomous robot navigation.Video surveillance in a dynamic environment, especially for humans and vehicles, is one of the current challenging research topics in computer vision. It is a key technology to fight against terrorism, crime, public safety and for efficient management of traffic. The work involves designing of the efficient video surveillance system in complex environments. In video surveillance, detection of moving objects from a video is important for object detection, target tracking, and behavior understanding. Detection of moving objects in video streams is the first relevant step of information and background subtraction is a very popular approach for foreground segmentation. In this thesis, we have simulated different background subtraction methods to overcome the problem of illumination variation, background clutter and shadows. Detecting and tracking of human body parts is important in understanding human activities. Intelligent and automated security surveillance systems have become an active research area in recent time due to an increasing demand for such systems in public areas such as airports, underground stations and mass events. In this context, tracking of stationary foreground regions is one of the most critical requirements for surveillance systems based on the tracking of abandoned or stolen objects or parked vehicles. Object tracking based techniques is the most popular choice to detect stationary foreground objects because they work reasonably well when the camera is stationary and the change in ambient lighting is gradual, and they also represent the most popular choice to separate foreground objects from the current frame. Surveillance networks are typically monitored by a few people, viewing several monitors displaying the camera feeds. It is very difficult for a human operator to effectively detect events as they happen. Recently computer vision research has to address ways to automatically some of this data, to assist human operators.

***Keywords*:** Object detection, Frame difference, Background subtraction, Gaussian mixture, Background modeling, Tracking, Block matching method, Kalman filter.

Contents

List of Figures

Chapter 1

Introduction

1.1 Object Detection and Tracking

Video surveillance is an active research topic in computer vision that tries to detect, recognize and track objects over a sequence of images and it also makes an attempt to understand and describe object behavior by replacing the aging old traditional method of monitoring cameras by human operators. Object detection and tracking are important and challenging tasks in many computer vision applications such as surveillance, vehicle navigation and autonomous robot navigation. Object detection involves locating objects in the frame of a video sequence. Every tracking method requires an object detection mechanism either in every frame or when the object first appears in the video. Object tracking is the process of locating an object or multiple objects over time using a camera. The high powered computers, the availability of high quality and inexpensive video cameras and the increasing need for automated video analysis has generated a great deal of interest in object tracking algorithms. There are three key steps in video analysis, detection interesting moving objects, tracking of such objects from each and every frame to frame, and analysis of object tracks to recognize their behavior.Therefore, the use of object tracking is pertinent in the tasks of, motion based recognition. Automatic detection, tracking, and counting of a variable number of objects are crucial tasks for a wide range of home, business, and industrial applications such as security, surveillance, management of access points, urban planning, traffic control, etc. However, these applications were not still playing an important part in consumer electronics. The main reason is that they need strong requirements to achieve satisfactory working conditions, specialized and expensive hardware, complex installations and setup procedures, and supervision of qualified workers. Some works have focused on developing automatic detection and tracking algorithms that minimizes the necessity of supervision. They typically use a moving object function that evaluates each hypothetical object configuration with the set of available detections without to explicitly compute their data association. Thus, a considerable saving in computational cost is achieved. In addition, the likelihood function has been designed to account for noisy, false and missing detections. The field of machine (computer) vision is concerned with problems that involve interfacing computers with their surrounding environment. One such problem, surveillance, has an objective to monitor a given environment and report the information about the observed activity that is of significant interest. In this respect, video surveillance usually utilizes electro-optical sensors (video cameras) to collect information from the environment. In a typical surveillance system, these video cameras are mounted in fixed positions or on pan-tilt devices and transmit video streams to a certain location, called monitoring room. Then, the received video streams are monitored on displays and traced by human operators. However, the human operators might face many issues, while they are monitoring these sensors. One problem is due to the fact that the operator must navigate through the cameras, as the suspicious object moves between the limited field of view of cameras and should not miss any other object while taking it. Thus, monitoring becomes more and more challenging, as the number of sensors in such a surveillance network increases. Therefore, surveillance systems must be automated to improve the performance and eliminate such operator errors. Ideally, an automated surveillance system should only require the objectives of an application, in which real time interpretation and robustness is needed. Then, the challenge is to provide robust and real-time performing surveillance systems at an affordable price. With the decrease in costs of hardware for sensing and computing, and the increase in the processor speeds, surveillance systems have become commercially available, and they are now applied to a number of different applications, such as traffic monitoring, airport and bank security, etc. However, machine vision algorithms (especially for single camera) are still severely affected by many shortcomings, like occlusions, shadows, weather conditions, etc. As these costs decrease almost on a daily basis, multi-camera networks that utilize 3D information are becoming more available. Although, the use of multiple cameras leads to better handling of these problems, compared to a single camera, unfortunately, multi-camera surveillance is still not the ultimate solution yet. There are some challenging problems within the surveillance algorithms, such as background modeling, feature extraction, tracking, occlusion handling and event recognition. Moreover, machine vision algorithms are still not robust enough to handle fully automated systems and many research studies on such improvements are still being done. This work focuses on developing a framework to detect moving objects and generate reliable tracks from surveillance video. The problem is most of the existing algorithms works on the gray scale video. But after converting the RGB vvideo frames to gray at the time of conversion, information loss occurs.The main problem comes when background and the foreground both have approximately same gray values. Then it is difficult for the algorithm to find out which pixel is foreground pixel and which one background pixel. Sometimes two different colors such as dark blue and dark violet, color when converted to gray scale, their gray values will come very near to each other,it can't be differentiated that which value comes from dark blue and which comes from dark violet. However, if color images are taken then the background and foreground color can be easily differentiated. So without losing the color information this modified background model will work directly on the color frames of the video.

1.2 Overview

In moving object detection various background subtraction techniques available in the literature were simulated. Background subtraction involves the absolute difference between the current image and the reference updated background over a period of time. A good background subtraction should be able to overcome the problem of varying illumination condition, background clutter, shadows, camouflage, bootstrapping and at the same time motion segmentation of foreground object should be done at the real time. It's hard to get all these problems solved in one background subtraction technique. So the idea was to simulate and evaluate their performance on various video data taken in complex situations.

Object tracking is a very challenging task in the presence of variability Illumination condition, background motion, complex object shape, partial and full object occlusions. Here in this thesis, modification is done to overcome the problem of illumination variation and background clutter such as fake motion due to the leaves of the trees, water flowing, or flag waving in the wind. Sometimes object tracking involves tracking of a single interested object and that is done using normalized correlation coefficient and updating the template.

On developing a framework to detect moving objects and generate reliable tracks from surveillance video. After setting up a basic system that can serve as a platform for further automatic tracking research, the question of variation in distances between the camera and the objects in different parts of the scene (object depth) in surveillance videos are takled. A feedback-based solution to automatically learn the distance variation in static-camera video scenes is implemented based on object motion in different parts of the scene. It gives more focus towards the investigation of detection and tracking of objects in video surveillance. The surveillance system is the process of monitoring the behavior, activities or other changing information, usually people for the purpose of influencing, managing, directing, and protecting. Most of the surveillance system includes static camera and fixed background which gives a clue for the object detection in videos by background subtraction technique. In surveillance system three main important steps these are object detection, object tracking and recognition. Some challenges in video processing Video analysis, video segmentation, video compression, video indexing. In case of video analysis there are three key steps: detection of interesting moving object, tracking of such objects from frame to frame and analysis of objects tracks to recognize their behavior. Next it comes video segmentation it means separation of objects from the background. It also consists of three important steps: object detection, object tracking and object recognition. In this work it is given more focus towards the investigation video analysis and video segmentation section.

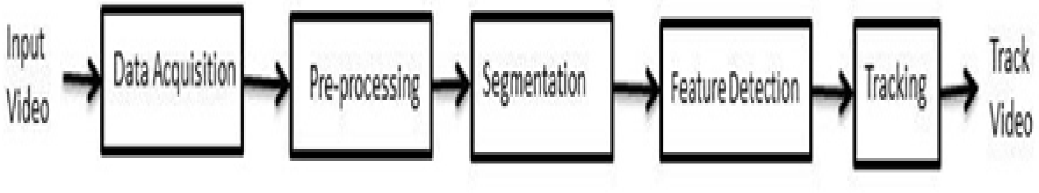

Figure 1.1: Analysis of detection and tracking approach

A typical automated single camera surveillance system usually consists of three main parts, which can be listed as moving object detection, object tracking and event recognition. In my problem it is to solve an automatic moving target detection and tracking details. The process of automatic tracking of objects begins with the identification of moving objects. An improved background subtraction method in conjunction with a novel yet simple background model to achieve very good segmentation is used. Once the moving pixels are identified, it is necessary to cluster these pixels into regions, which is referred as blobs, so that pixels belonging to a single object are grouped together. Single moving objects are often incorrectly separated into two or more sub regions because of lack of connectivity between pixels, which usually occurs due to occlusion from other objects.

1.3 Motivation

After studying the literature, it is found that detecting the object from the video sequence and also track the object it is a really challenging task. Object tracking can be a time consuming process due to amount of data that is contained in the video. From the literature survey it is found that there are many background subtraction algorithm exits which work efficiently in both indoor and outdoor surveillance system. Julio et al. [3] has proposed a background modeling technique and used another algorithm to detect shadowed region. But the shadow removal technique is an overhead for object tracking algorithm. It will be better if the shadow can be removed at the time of the foreground object detection algorithm by designing an efficient algorithm, which can properly classify the foreground object and background removing false foreground pixel from detection. Then there will no extra computation needed for shadow detection and removal.

Video surveillance is the most active research topic in computer vision for humans and vehicles. Here the aim is to develop an intelligent visual surveillance system by re-placing the age old tradition method of monitoring by human operators. The motivation in doing is to design a video surveillance system for motion detection, and object tracking.

The area of automated surveillance systems is currently of immense interest due to its implications in the field of security. Surveillance of vehicular traffic and human activities offers a context for the extraction of significant information such as scene motion and traffic statistics, object classification, human identification, anomaly detection, as well as the analysis of interactions between vehicles, between humans or between vehicles and humans. A wide range of research possibilities is open in relation to video surveillance and tracking.

1.4 Objective

This thesis aims to improve the performance of object detection and tracking by contributing originally to two components (a) motion segmentation (b) object tracking.

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. Tracking is not a trivial task given the non-deterministic nature of the subjects, their motion, and the image capture process itself. The objective of video tracking is to associate target objects in consecutive video frames. The association can be especially difficult when the objects are moving fast relative to the frame rate.

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 goal of the work in this thesis is twofold:

1. (a) To set up a system for automatic segmentation and tracking of moving objects in stationary camera video scenes, which may serve as a foundation for higher level reasoning tasks and applications

2. (b) To make significant improvements in commonly used algorithms. Finally, the aim is to show how to perform detection and motion-based tracking of moving objects in a video from a stationary camera.

Therefore the main objectives are:

- To analyze segmentation algorithm to detect the objects.

- To analyze some tracking method for tracking the single objects and multiple objects.

1.5 Thesis Organization

The rest of the thesis is organized as follows:

Chapter2: This chapter discusses about the background concepts related to this project work. The chapter also discusses object segmentation in image sequences, background modeling and tracking approaches. The architecture and block diagram of tracking flow systems are also explained in this chapter.

Chapter 3: The literature surveys that have been done during the research work has ben discussed here. It also provides a detailed survey of the literature related to motion detection and object tracking.Discussion about the existing and some new methods for detection and tracking of objects are done. In this chapter existing methods are discussed and also examined. This chapter presents the methodology and implementation of some existing and experimental results subsequently

Chapter 4: This chapter provides concluding comments those can be made to the project.

Chapter 2

Background

2.1 Introduction

Object tracking is an important job within the field of computer vision. Object detection involves locating objects in frames of a video sequence. Tracking is the process of locating moving objects or multiple objects over a period of time using a camera. Technically, tracking is the problem of estimating the trajectory or path of an object in the image plane as it moves around a scene. The high-powered computers, the availability of high quality and inexpensive video cameras, and the increasing need for automated video analysis has generated a great deal of interest in object tracking algorithms. There are three key steps in video analysis:

- Detection of interesting moving objects.
- Tracking of such objects from frame to frame.
- Analysis of object tracks to recognize their behavior.

So now the question arises here that, where object tracking is suitable to apply? Mainly the use of object Tracking is pertinent in the task of:

- Motion-based recognition

- automated surveillance
- video indexing
- human-computer interaction
- traffic monitoring
- vehicle navigation

Tracker assigns consistent labels to the tracked objects in different frames of a video. Additionally, depending on the tracking domain, a tracker can also provide object-centric information, such as orientation, area or shape of an object. Tracking objects can be complex due to:

- Loss of information caused by projection of the 3D world on a 2D image,
- Noise in images,
- complex object motion,
- non rigid or articulated nature of objects,
- partial and full object occlusions,
- complex object shapes,
- scene illumination changes, and
- Real-time processing requirements.

  By imposing constraints on the motion and appearance, objects can be tracked. Almost all tracking algorithms assume that the object motion is smooth with no abrupt changes. One can constrain the object motion to be of constant velocity or a constant acceleration based on prior information. Huge knowledge about the number and the size of objects, or the object appearance and shape, can also be used to simplify the problem. A number of approaches for object tracking have been proposed. These differ from each other based on the way they approach the following questions:

- Which object representation is suitable for tracking?
- Which image features should be used?
- How should the motion, appearance and shape of the object be modeled?

The answers to these questions depend on the environment in which the tracking is performed and the end use for which the tracking information is being sought. A large number of tracking methods have been proposed which to answer these questions.

2.2 Object Representation

An object is simply nothing but an entity of interest. Objects can be represented by their shapes and appearances. For example, boats on the sea, fish in an aquarium, vehicles on a road, planes in the air, people walking on a road may be important to track in a specific domain. So there are various representations of object shape, which is commonly used for tracking and then addresses the joint shape and appearance representations in describing [1] as follows.

- Points-The object is represented by a point, which is the centroid (Figure 2 (a)) or a set of points (figure 2 (b)). The point representation is suitable when it is given more concentration on the object which occupied small regions in an image.
- Primitive geometric shape-Geometric shape i.e. The object shape is represented by rectangle, ellipse (Figure 2 (c), (d)). Primitive geometric shapes are more suitable for representing simple rigid objects as well as non-rigid objects.

- Object silhouette and contour-Contour is the boundary of an object (Figure 2 (g), (h)). The region inside the contour is called the silhouette of the object (Figure 2 (I)). These representations are suitable for tracking complex non rigid shapes.

- Articulated shape models-Articulated objects are composed of body parts that are held together by joints (Figure 2 (e)). For example, the human body is an articulated object with torso, legs, hands, head, and feet connected by joints. The relationship between the parts is governed by kinematic motion models, for example, joint angle, etc.

- Skeletal model- Object skeleton can be extracted by applying medial axis transform to the object silhouette. This model is commonly used as a shape representation for recognizing objects. Skeleton representation can be used to model both articulated and rigid objects (see figure 2 (f))

Object representations. (a) Centroid, (b) multiple points, (c) rectangular patch, (d) elliptical patch, (e) part-based multiple patches, (f) object skeleton, (g) complete object contour, (h) control points on object contour, (i) object silhouette.

Similarly there are a various ways to represent the appearance feature of objects. It should be noted that the shape representation can be combined with appearance representations for tracking. Some common appearance representations in the case of object tracking are described in [1] as follows.

- Probability densities of object appearance-The probability density estimates the object appearance can either be parameters, such as Gaussian and a mixture of Gaussians , such as Parzen windows and histograms. The probability densities of object appearance features (color, texture) can be computed from the image regions specified by the shape models (interior region of an ellipse or a contour) [1].

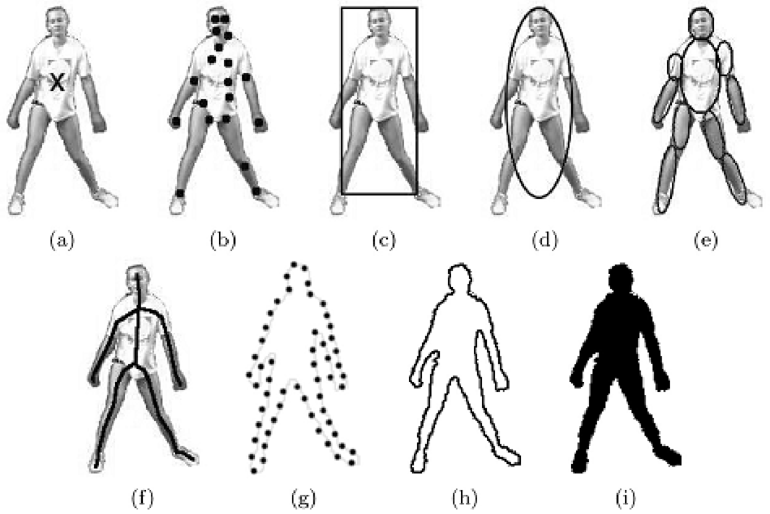

Figure 2.1: Object Representation[1]

- Templates-Templates are formed using simple geometric shapes or silhouettes. It carries both spatial and appearance information. Templates, however, only encode the object appearance generated from a single view. Thus, they are only suitable for tracking objects whose poses do not vary considerably during the course of tracking.

- Active appearance models-these are generated by simultaneously modeling the object shape and appearance. Object shape is defined by a set of landmarks. Each landmark, an appearance vector is stored in the form of color, texture or gradient magnitude. These models required a training phase where both the shapes & its associated appearance is learned from a set of samples.

- Multiview appearance models- These models encode different views of an object. One approach to represent the different object views is to generate a subspace from the given view. Example of subspace approaches is Principal Component Analysis (PCA), Independent component Analysis (ICA). One limitation of multi-view appearance models is that the appearances in all views have required a lot of time.

In general, there is a strong relationship between object representation and tracking algorithms. Object representations are chosen according to the application domain.

2.3 Object Detection

Every tracking method requires an object detection mechanism either in every frame or when the object first appears in the video. A common approach for object detection is to use information in a single frame. However, some object detection methods make use of the temporal information computed from a sequence of frames to reduce the number of false detections. For object detection, there are several common object detection methods described in [1].

1. Point detectors-Point detectors are used to find interesting points in images which have an expressive texture in their respective localities. A desirable quality of an interest point is its invariance to changes in illumination and camera viewpoint. In literature, commonly used interest point detectors include Moravec's detector, Harris detector, KLT detector, SIFT detector.

2. Background Subtraction-Object detection can be achieved by building a representation of the scene called the background model and then finding deviations from the model for each incoming frame. Any significant change in an image region from the background model signifies a moving object. The pixels constituting the regions undergoing change are marked for further processing. This process is referred to as the background subtraction. There

are various methods of background subtraction as discussed in the survey [1] are Frame differencing Region-based (or) spatial information, Hidden Markov models (HMM) and Eigen space decomposition.

3. Segmentation-The aim of image segmentation algorithms is to partition the image into perceptually similar regions. Every segmentation algorithm addresses two problems, the criteria for a good partition and the method for achieving efficient partitioning. In the literature survey it has been discussed various segmentation techniques that are relevant to object tracking [1] They are, mean shift clustering, and image segmentation using Graph-Cuts (Normalized cuts) and Active contours.

Object detection can be performed by learning different object views automatically from a set of examples by means of supervised learning mechanism.

2.4 Object Tracking

The aim of an object tracker is to generate the trajectory of an object over time by locating its position in every frame of the video [1]. But tracking has two definition one is in literally it is locating a moving object or multiple object over a period of time using a camera. Another one in technically tracking is the problem of estimating the trajectory or path of an object in the image plane as it moves around a scene. The tasks of detecting the object and establishing a correspondence between the object instances across frames can either be performed separately or jointly. In the first case, possible object region in every frame is obtained by means of an object detection algorithm, and then the tracker corresponds objects across frames. In the latter case, the object region and correspondence is jointly estimated by iteratively updating object location and region information obtained from previous frames [1]. There are different methods of Tracking.

- Point is tracking- Tracking can be formulated as the correspondence of detecting objects represented by points across frames. Point tracking can be divided into two broad categories, i.e. Deterministic approach and Statistical approach. Objects detected in consecutive frames are represented by points, and the association of the points is based on the previous object state which can include object position and motion.

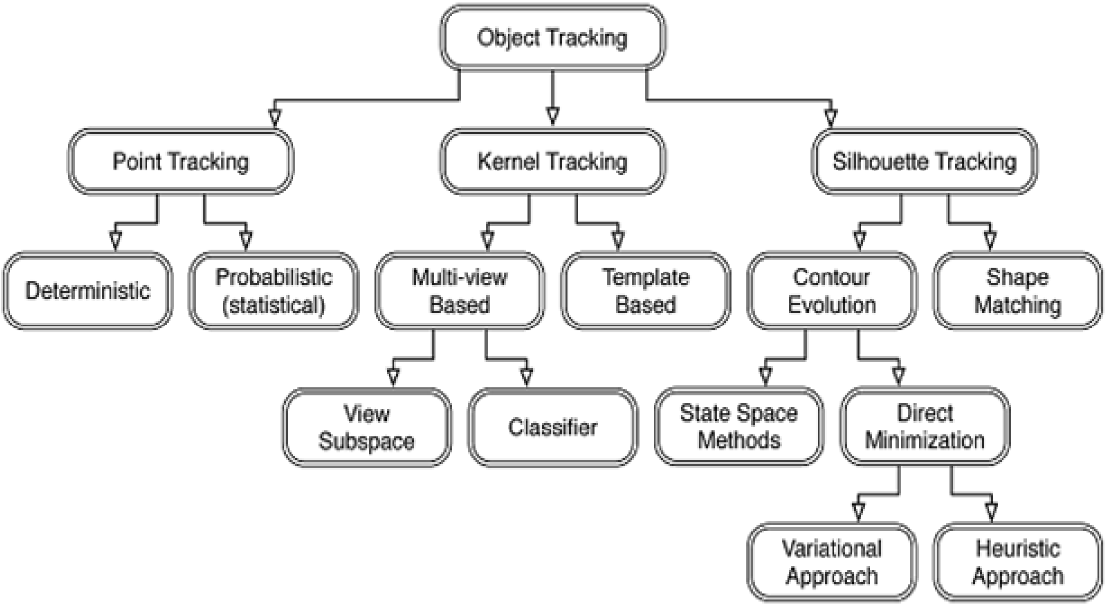

Figure 2.2: Different Tracking Categories [1]

- Kernel tracking- Performed by computing the motion of the object, represented by a primitive object region, from one frame to the next. Object motion is in the form of parametric motion or the dense flow field computed in subsequent frames. Kernel tracking methods are divided into two subcategories based on the appearance representation used i.e. Template and Density-based Appearance Model and Multi-view appearance model.

- Silhouette Tracking- It Provides an accurate shape description of the target objects. The goal of silhouette tracker is to find the object region in each frame by means of an object model generated using the previous frames. Silhouette trackers can be divided into two categories i.e. Shape matching and Contour tracking.

Object tracking consists in estimating of the trajectory of moving objects in the sequence of images. The most important is the automation of object tracking is a challenging task. Dynamics of multiple parameters, changes representing features and motion of the objects and temporary partial or full occlusion of the tracked objects have to be considered.

2.4.1 Feature Selection for Tracking

It plays a vital role to select a proper feature in tracking. So feature selection is closely related to the object representation. For example, color is used as a feature for histogram based appearance representations, while for contour-based representation, object edges is usually used as features. Generally, many tracking algorithms use a combination of these features. The details of common visual features are as follows :

- Color-Color of an object is influenced by two factors. They are Spectral power distribution of the illuminant and Surface reflectance properties of the object. Different color models are RGB, L*u*v and L*a*b used to represent color.

- Edges-Edge detection is used to identify strong changes in image intensities generated by object boundary. Edges are less sensitive to illumination changes compared to color features. Most popular edge detection approach is Canny Edge detector.

- Optical Flow-It is a dense field of displacement vector which defines the translation of each pixel in a region. It is computed using the brightness constraint, which assumes brightness constancy of corresponding pixels in consecutive frames. Optical Flow is commonly used as a feature in motion based segmentation and tracking application.

- Texture-Texture is a measure of the intensity variation of a surface which quantifies properties such as smoothness and regularity. It requires a processing step to generate the descriptors.There are various texture descriptors: Gray-Level Co-occurrence Matrices, loss texture measures, wavelets, and steerable pyramids.

Mostly features are chosen manually by the user depending on the application. The problem of automatic feature selection has received significant attention in the pattern recognition community. Automatic feature selection methods can be divided into, Filter Methods and Wrapper Methods. Filter methods try to select the features based on general criteria, whereas Wrapper methods selects the features based on the usefulness of the features in a specific problem domain.

2.4.2 Single Camera Object Tracking:

Till now various concepts of object tracking are being disscussed.It is necessary to know how the tracking occurs in front of a single fix camera. It is very important to track properly in a single camera so that it will be easy for us to track it in multiple cameras. Whenever tracking is being done using single camera there are various challenges need to be taken care of.

- A single person may enter in to the FOV of camera.

- More than one person may enter in to the FOV of camera.

- Object carrying some goods.

First Motion Segmentation to extract moving blobs in the current frame is performed. Some blobs that are very small and are likely to be noise are deleted. The object tracking module tracks these moving objects over successive frames to generate object tracks.

2.4.3 Segmentation

After preprocessing the next step is segmentation.Segmentation means, separate the objects from the background.The aim of image segmentation algorithms is to partition the image in to perceptually similar regions.Every segmentation algorithm addresses two problems, the criteria for a good partition and the method for achieving efficient partitioning. In the literature survey it has been discussed various segmentation techniques that are relevant to object tracking [1]. They are, Mean shift clustering, and image segmentation using Graph-cuts (Normalized cuts) and Active contours.

2.4.4 Motion Segmentation

The first job in any surveillance application is to distinguish the target objects in the video frame. Most pixels in the frame belong to the background and static regions, and suitable algorithms are needed to detect individual targets in the scene. Since motion is the key indicator of target presence in surveillance videos, motion-based segmentation schemes are widely used. An effective and simple method for approximating the background that enables the detection of individual moving objects in video frames is being utilized.

2.4.5 Segmentation Methods

These are the different segmentation technique which will be discussed in details in chapter 4.There are numerous proposals for the solution of moving object detection problem in the surveillance system. Although there are numerous proposals for the solution of moving object detection problem in surveillance systems, some of these methods are found out to be more promising by the researchers in the field. Methods are like

- Frame differencing method to detect objects.
- Mixture of Gaussian based on moving object detection method.
- Background subtraction method to detect foreground objects.

2.4.6 Foreground Segmentation

Foreground segmentation is the process of dividing a scene into two classes; one is foregrounding another one is background. The background is the region such as roads, buildings and furniture. While the background is fixed, its appearance can be expected to change over time due to factors such as changing weather or lighting conditions. The foreground any element of the scene that is moving or expected to move and some foreground elements may actually be stationary for long periods of time such as parked cars, which may be stationary for hours at a time. It is also possible that some elements of background may actually move, such as trees moving in a breeze.

The main approaches to locating foreground objects within in the surveillance system is

1. Background modeling or subtraction-incoming pixels compare to a background model to determine if they are foreground or background.

2.4.7 Moving object Detection

The performance of a surveillance system considerably depends on its first step that is detection of the foreground objects which do not belong to the background scene. These foreground regions have a significant role in subsequent actions, such as tracking and event detection.

The objective of video tracking is to associate target objects in consecutive video frames. The association can be especially difficult when the objects are moving fast relative to the frame rate. Another situation that increases the complexity of the problem is when tracked object changes orientation over time. For these situations video tracking systems usually employ a motion model which describes how the image of the target might change for the different possible motion of the objects. Numerous approaches for object tracking have been proposed. These primarily differ from each other based on the way they approach the following questions: Which object representation is suitable for tracking? Which image features should be used? How should the motion, appearance and shape of the object be modeled?

The answers to these questions depend on the context or environment in which the tracking is performed and the end use for which the tracking information is being sought.

Moving object segmentation is simply based on a comparison between the input frame and a certain background mode, and different regions between the input and the model are labeled as foreground based on this comparison. This assessment can be the simple frame differencing, if the background is static (has no moving parts and is easier to model). However, more complex comparison methods are required to segment foreground regions when background scenes have dynamic parts, such as moving tree branches and bushes. In the literature, there are various algorithms, which can cope with these situations that will be discussed in the following sections [2].

Nearly, every tracking system starts with motion detection. Motion detection aims at separating the corresponding moving object region from the background image. The first process in the motion detection is capturing the image information using a video camera. The motion detection stage includes some image preprocessing step such as; gray-scaling and smoothing, reducing image resolution using low resolution image technique, frame difference, morphological operation and labeling. The preprocessing steps are applied to reduce the image noise in order to achieve a higher accuracy of the tracking. The smoothing technique is performed by using median filter. The lower resolution image is performed in three successive frames to remove the small or fake motion in the background. Then the frame difference is performed on those frames to detect the moving object emerging on the scene. The next process is applying a morphological operation such as dilation and erosion as filtering to reduce the noise that remains in the moving object. Connected component labeling is then performed to label each moving object in different label.

The second stage is tracking the moving object. In this stage, a block matching technique to track only the interest moving object among the moving objects emerging in the background, is performed. The blocks are defined by dividing the image frame into non-overlapping square parts. The blocks are made based on PISC image that considers the brightness change in all the pixels of the blocks relative to the considered pixel.

The last stage is object identification. For this purpose spatial and color information of the tracked object as the image feature is used. Then, a feature queue is created to save the features of the moving objects. When the new objects appear on the scene, they will be tracked and labeled, and the features of the object are extracted and recorded into the queue. Once a moving object is detected, the system will extract the features of the object and identify it from the identified objects in the queue. A few details of each stage are described as follows.

2.4.8 Object Detection

Performance of an automated visual surveillance system considerably depends on its ability to detect moving objects in the observed environment. A subsequent action, such as tracking, analyzing the motion or identifying objects, requires an accurate

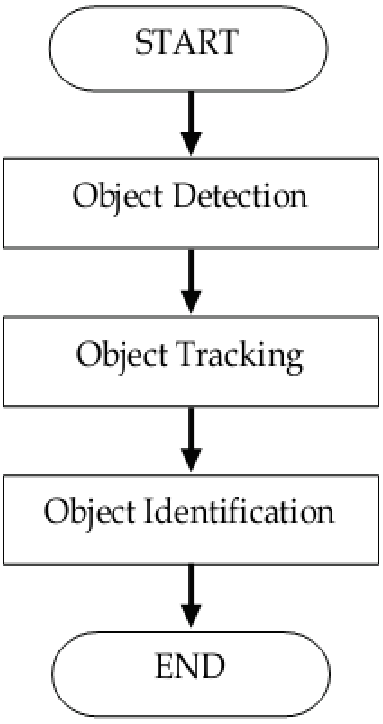

Figure 2.3: Flow Of Procedure[2]

extraction of the foreground objects, making moving object detection a crucial part of the system. In order to decide on whether some regions in a frame are foreground or not there should be a model for the background intensities. Any change, which is caused by a new object, should be detected by this model, whereas un-stationary background regions, such as branches and leaves of a tree or a flag waving in the wind, should be identified as a part of the background. So to handle these problems a method was proposed.

Mainly object detection method consists of two main steps. The first step is a preprocessing step including gray scaling, smoothing, and reducing image resolution and so on. The second step is filtering to remove the image noise contained in the object. The filtering is performed by applying the morphology filter such as dilation and erosion. And finally connected component labeling is performed on the filtered image

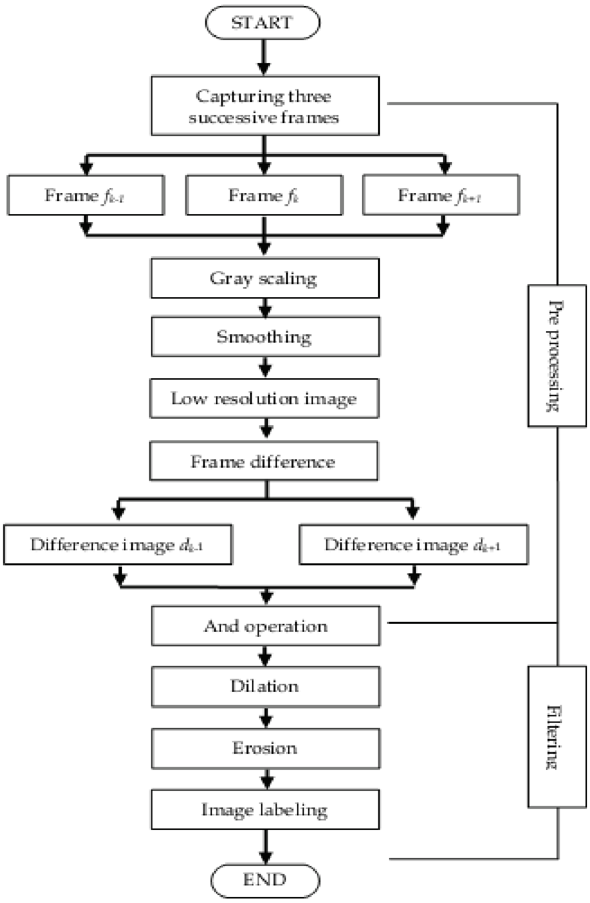

Figure 2.4: Flow Of Object Detection[2]

Pre-processing : In the preprocessing phase, the first step of the moving object detection process is capturing the image information using a video camera. In order to reduce the processing time, a grayscale image is used on entire process instead of the color image. The grayscale image only has one color channel that consists of 8 bits while RGB image has three color channels. Image smoothing is performed to reduce image noise from input image in order to achieve high accuracy for detecting the moving objects. The smoothing process is performed by using a median filter with $m \times m$ pixels. Here, un-stationary background such as branches and leaf of a tree as part of the background are considered. The un-stationary background often considers as a fake motion other than the motion of the object interest and can cause the failure of detection of the object. To handle this problem, the resolution of the image is reduced to be a low resolution image. A low resolution image is done by reducing spatial resolution of the image with keeping the image size. The low resolution image can be used for reducing the scattering noise and the small fake motion in the background because of the un-stationary background such as leaf of a tree. These noises that have small motion region will be disappeared in low resolution image.

Next it comes filtering phase. In order to fuses narrow breaks and long thin gulfs, eliminates small holes, and fills gaps in the contour, a morphological operation is applied to the image. As a result, small gaps between the isolated segments are erased and the regions are merged. To extract the bounding boxes of detecting objects, connected component analysis was used. Morphological operation is performed to fill small gaps inside the moving object and to reduce the noise remained in the moving objects. The morphological operators implemented are dilation followed by erosion. In dilation, each background pixel that is touching an object pixel is changed into an object pixel. Dilation adds pixels to the boundary of the object and closes isolated background pixel. Dilation can be expressed as:

$$f(x,y) = \begin{cases} 1, & if\, there\, is\, one\, or\, more\, pixels\;\, of\, the\, 8\, neighbors\, are\, 1 \\ 0, & otherwise \end{cases} \tag{2.1}$$

In erosion, each object pixel that is touching a background pixel is changed into a background pixel. Erosion removes isolated foreground pixels. Erosion can be expressed as:

$$f(x,y) = \begin{cases} 0, & if\, there\, is\, one\, or\, more\, pixels\;\, of\, the\, 8\, neighbors\, are\, 0 \\ 1, & otherwise \end{cases} \tag{2.2}$$

2.4.9 Tracking Methods:

- Point is tracking- Tracking can be formulated as the correspondence of detecting objects represented by points across frames. Point tracking can be divided into two broad categories, i.e. Deterministic approach and Statistical approach.

- Kernel tracking- Performed by computing the motion of the object, represented by a primitive object region, from one frame to the next. Object motion is in the form of parametric motion or the dense flow field computed in subsequent frames. Kernel tracking methods are divided into two subcategories based on the appearance representation used i.e. Template and Density-based Appearance Model and Multi-view appearance model.

- Silhouette Tracking-It Provides an accurate shape description of the target objects. The goal of silhouette tracker is to find the object region in each frame by means of an object model generated using the previous frames. Silhouette trackers can be divided into two categories i.e. Shape matching and Contour tracking.

2.4.10 Prediction Methods:

An important part of a tracking system is the ability to predict where an object will be next frame. This is needed to aid the matching the tracks to detect objects and to predict the position during occlusion. There are four common approaches to predict the objects' positions:

1. Block matching

2. Kalman filters

3. Motion models

4. particle filter

2.5 Object Tracking

After the object detection is achieved, the problem of establishing a correspondence between object masks in consecutive frames should arise. Obtaining the correct track information is crucial for subsequent actions, such as object identification and activity recognition. A block matching technique is used for this purpose.

2.5.1 Block Matching Method

The entire process of tracking the moving object is illustrated in the following Fig 2.6. The block matching method is well described in [4], which is applied here.

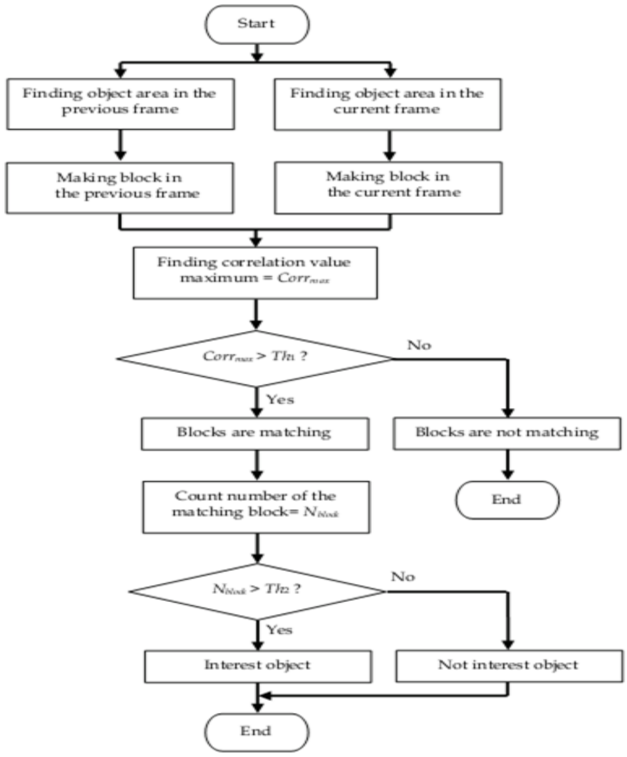
Start
Finding object area in the previous frame
Finding object area in the current frame
Making block in the previous frame
Making block in the current frame
Finding correlation value maximum = $Corr_{max}$
$Corr_{max} > Th_1$?
No
Yes
Blocks are matching
Blocks are not matching
Count number of the matching block= N_{block}
End
$N_{block} > Th_2$?
No
Yes
Interest object
Not interest object
End

Figure 2.5: Flow of tracking interest object[2]

Block matching is a technique for tracking the interest moving object among the moving objects emerging on the scene. In this article, the blocks are defined by dividing the image frame into non-overlapping square parts. The blocks are made based on peripheral increment sign correlation (PISC) image that considers the brightness change in all the pixels of the blocks relative to the considered pixel. In the Fig it shows the block in the PISC image with block size is 55 pixels. Therefore, one block consists of 25 pixels. The blocks of the PISC image in the previous frame are defined as shown in Eq. (2.3). Similarly, the blocks of the PISC image in the current frame are defined in Eq. (2.4). To determine the matching criteria of the blocks in two successive frames, the evaluation is done using a correlation value that expresses in Eq. (2.5). This equation calculates the correlation value between block in the previous frame and the current one for all pixels in the block. The high correlation value shows that the blocks are matching each other. The interest moving object is determined when the number of matching blocks in the previous and current frame are higher than the certain threshold value. The threshold value is obtained experimentally.

$$b_{np} = \begin{cases} 1, & iff_{np \geq f(i,j)} \\ o, & otherwise \end{cases} \tag{2.3}$$

$$b'_{np} = \begin{cases} 1, & iff_{np \geq f(i,j)} \\ o, & otherwise \end{cases} \tag{2.4}$$

$$corr^{n} = \sum_{P=0}^{N} b_{np} * b'_{np} + \sum_{P=0}^{N} (1 - b_{np}) * (1 - b'_{np}) \tag{2.5}$$

where p & p' are the block in the previous and current frame, n is the number of block and N is the number of pixels of block.

2.5.2 Tracking Method

The tracking method used in this article can be described as follows. The matching process is illustrated in Fig.2.6. Firstly, blocks and the tracking area are made only in the area of moving objects to reduce the processing time.The block size (block A) is made with 9x9 pixels in the previous frame. It is assumed that the object coming firstly will be tracked as the interest moving object. The block A will search the matching block in each block of the current frame by using correlation value as expressed in Eq. (2.5). In the current frame, the interest moving object is tracked when the object has maximum number of matching blocks. When that matching criteria are not satisfied, the matching process is repeated by enlarging the tracking area (the rectangle with dash line). The blacks still are made in the area of moving objects. When the interest moving object still cannot be tracked, then the moving object is categorized as not interest moving object or another object and the tracking process is begun again from the begin.

2.5.3 Feature Extraction

The feature of objects extracted in the spatial domain is the position of the tracked object. The spatial information combined with the features in time domain represents the trajectory of the tracked object, so the movement and the speed of the moving objects that are tracked can be estimated. Therefore, the features of spatial domain are very important to object identification. The bounding box defined in Eq. (2.4) is used as spatial information of moving objects.

After getting the interest moving object, that is extracted by using a bounding box. The bounding box can be determined by computing the maximum and minimum value of x and y coordinates of the interest moving object according to the following equation:

$$B^i_{min} = \left\{(x^i_{min}, y^i_{min}) | x, y \in O^i \right\} \tag{2.6}$$

$$B^i_{max} = \left\{(x^i_{max}, y^i_{max}) | x, y \in O^i \right\} \tag{2.7}$$

where O^i denotes set of coordinate of points in the interest moving object i,B^i_{min} is the left top corner cordinates of the interest moving object i, and B^i_{max} is the right bottom corner cordinates of the interesting moving object i. In the chapter 4 shows the bounding box of the object tracking.

2.6 Kalman Filter

A Kalman filter is used to estimate the state of a linear system where the state is assumed to the distributed by a Gaussian. The Kalman filter is a recursive predictive filter that is based on the use of state space techniques and recursive algorithms. It is estimated the state of a dynamic system. This dynamic system can be disturbed by some noise, mostly assumed as white noise. To improve the estimated state the Kalman filter uses measurements that are related to the state but disturbed as well. Kalman filtering is composed of two steps. Thus the Kalman filter consists of two steps:

- The prediction
- The correction

In the first step the state is predicted with the dynamic model. The prediction step uses the state model to predict the new state of the variables.

$$\overline{X}^t = DX^{t-1} + W \tag{2.8}$$

$$\overline{\sum}^t = D\sum\limits^{t-1} D^{^t} + Q^{^t} \tag{2.9}$$

Where X^t and $\sum^t$ are the state and covariance predictions at time t. D is the state transition matrix which defines the relation between the state variables at time t and t-1. Q is the covariance of the noise W. Similarly the correction step uses the current observation Z^t to update the object state

$$K^t = \sum\limits^{t} M^t [M \sum\limits^{t} M^t + R^t]^{-1} \tag{2.10}$$

$$X^t = \overline{X}^t + K^t[Z^t - MX^t] \tag{2.11}$$

$$\sum\limits^{t} = \sum\limits^{t} - K^t M \sum\limits^{t} \tag{2.12}$$

where M is the measurement matrix, K is the Kalman gain which is called as the Riccati equation used for propagation of the state models. The updated state

X^t is distributed by Gaussian. Similarly Kalman filter and extended Kalman filter assumes that the state is distributed by a Gaussian.

In the second step it is corrected with the observation model, so that the error covariance of the estimator is minimized. In this sense it is an optimal estimator. Kalman filter has been extensively used in the vision community for tracking .

2.6.1 General Application

The basic components of the Kalman filter are the state vector, the dynamic model and the observation model, which are described below

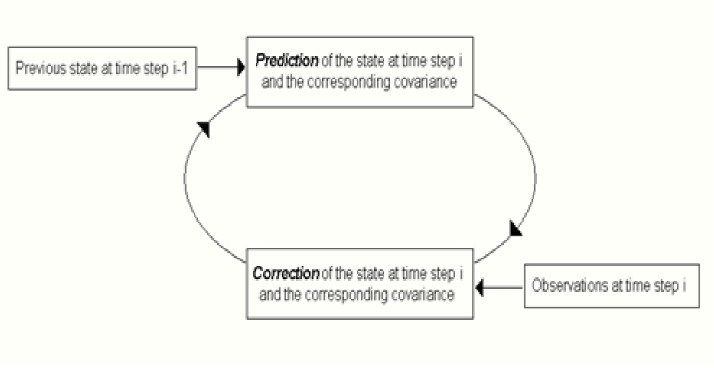
Previous state at time step i-1
Prediction of the state at time step i and the corresponding covariance
Correction of the state at time step i and the corresponding covariance
Observations at time step i

Figure 2.6: Basic Components of Filter[34]

2.6.2 State Vector

The state vector contains the variables of interest. It describes the state of the dynamic system and represents its degrees of freedom. The variables in the state vector cannot be measured directly but they can be inferred from the values that are measurable. Elements of the state vector can be positioned, velocity, orientation angles, etc. A very simple example is a train that is driving with a constant velocity on a straight rail. In this case the train has two degrees of freedom, the distance and the velocity. The state vector has two values at the same time; one is the predicted value before the update and the posterior value after the update.

2.6.3 The Discrete Kalman Filter

In 1960, R.E. Kalman published his famous paper describing a recursive solution to the discrete data linear filtering problem. Since that time, due in large part to advances in digital computing; the Kalman filter has been the subject of extensive research and application, particularly in the area of autonomous or assisted navigation.

The main problem with Kalman filtering is that statistical models are required for the system and the measurement instruments. Unfortunately, they are typically not available, or difficult to obtain. The two most commonly recommended methods of approaching this problem are:

- Employ an adaptive algorithm which adjusts these unknown parameters (such as the measurement noise variance) after each time step based on the observed measurements. This also accounts for processes with changing parameters.

- Perform an ”online” analysis of the system and measurement instruments prior to running the process (system identification). It should be noted however that the second approach will not always be applicable if the process cannot be observed directly. In other words, if the measurements in the online analysis also contain errors, the process cannot be accurately profiled.

Each x_t contains an $m \times 1$ mean vector $\wedge X$ and an $m \times n$ covariance matrix P, where m is the number of parameters that describe the state. A simple example of the parameters necessary for tracking are the x and y coordinates as well as the u and v velocity components. The y_t nodes are represented by an $n \times 1$ vector which is nothing but the observed position of the target in the context of video tracking. This method of describing the state through a finite set of parameters is known as the state-space model (SSM).

As mentioned earlier, the state nodes are related to each other through the physics underlying object motion. The transition from one state to the next could be described in many ways. These different alternatives can be grouped into linear and nonlinear functions describing the state transition. Although it is possible to handle either of these transition types, the standard Kalman filter employs a linear transition function [2].

The extended Kalman filter (EKF) allows a non-linear transition, together with a non-linear measurement relationship. For the standard Kalman filter, the state transition from t to t + 1 can be expressed by the equation

$$X_{t+1} = Ax_t + W_t \tag{2.13}$$

Where A is referred to as the state transition matrix and w is a noise term. This noise term is a Gaussian random variable with zero mean and a covariance matrix Q, so its probability distribution is

$p\left(w\right) \sim N\left(O, Q\right)$

The covariance matrix Q will be referred to as the process noise covariance matrix in the remainder of this report. It accounts for possible changes in the process between t and t + 1 that are not already accounted for in the state transition matrix. Another assumed property of w is that it is independent of the state x_t. The measurement is taken, the node y becomes observed and x node can be

$$y_t = Cx_t + w_t \tag{2.14}$$

Where C is an $m \times n$ matrix which relates the state to the measurement. Much like wt. Now that a graphical model has been established and the relationships between the nodes are formulated, it is possible to look at how these relationships can be used in the tracking process. A prediction is required at each time step before the target is located with the tracking algorithm. The predicted position is nothing but the expected measurement given all the previous measurements. It illustrates the situation before each prediction is made and serves as a template for the recursive step. Initially Y_t is observed, and a prediction for Y_{t+1} is required. After the prediction is made and measurement is taken, y becomes observed, and the process t+1 repeats for t + 2 [2].

Tracking more than just the x and y coordinates of the target would be interesting and is possible. It would be especially interesting to study the applicability of the predication methods to these further degrees of freedom such as scaling or rotation.

When thinking along these lines, why not focus on a single dimension rather than using a 2D image and allowing both x and y translations? This approach was considered, and would include the analysis of one-dimensional intensity vectors rather than 2D images. It is in fact expected to provide clearer results. However, to study the effects of prediction on real image sequences, a 2D implementation was required. Furthermore, even though having results in an abstract setting is useful; a model with features a little more similar to real image sequences was desired. On average, the results achieved with the Kalman filter should be at least as good as those of the simple prediction method. This is expected to be helpful in the experiments to measure the performance of Kalman filter.

2.6.4 Discrete Kalman Filter Algorithm

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. As such, the equations for the Kalman filter fall into two groups:

- Time update equations
- Measurement update equations

The time update equations are responsible for projecting forward (in time) the current state and error covariance estimates to obtain the next time step. The measurement update equations are responsible for the feedback-i.e. for incorporating a new measurement into the improved estimate. The time update equations can also be thought of as predictor equations, while the measurement update equations can be thought of as corrector equations. The specific equations for the time and measurement updates are presented below

Discrete Kalman filter time update equations.

$$X_k = Ax_{k-1} + Bu_{k-1} \tag{2.15}$$

$$P_k = AP_{k-1}A^{T} + Q \tag{2.16}$$

Discrete Kalman filter measurements update equations

$$K_k = P_k^{-} H^T + P_k H + R^{-} \tag{2.17}$$

$$X_k = X_k^{-} - H^{\wedge} x_k \tag{2.18}$$

This recursive nature is one of the very appealing features of the Kalman filter it makes practical implementations much more feasible than an implementation of a which is designed to operate on all of the data directly for each estimate. The Kalman filter instead recursively conditions the current estimate on all of the past measurements.

2.6.5 Algorithm Discussion

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 measurements.

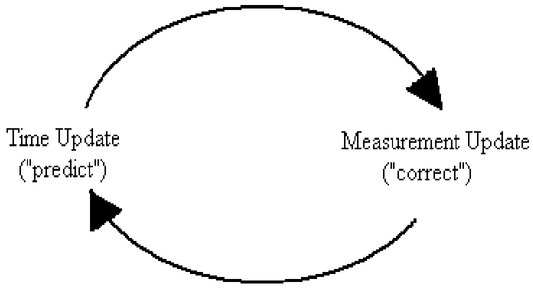

Figure 2.7: The discrete kalman Filter cycle [2]

As such, 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 posterior estimate. The time update equations can also be thought of as predictor equations, while the measurement update equations can be thought of as corrector equations. Indeed the final estimation algorithm resembles that of a predictor-corrector algorithm as shown in figure.

The specific equations for the time and measurement updates are presented below.

Discrte Kalman Filter Time Update

$$\widehat{x}_k^- = Ax_{K-1} + Bu_{k-1} \tag{2.19}$$

$$P_k^- = AP_{k-1}A^T + Q \tag{2.20}$$

As shown the time update equations above project the state and covariance estimates forward from time step to step.

Discrete Kalman Filter Measurement Update

$$K_k = P_k^- H^T(HP_k^- H^T + R)^{-1} \tag{2.21}$$

$$\widehat{x}_k{}^{=}\widehat{x}_k{}^{-} + K(Z_k - H)\widehat{x}_k{}^{-} \tag{2.22}$$

$$P_k = (I - K_k H)P_k^{-1} \tag{2.23}$$

The first task during the measurement update is to compute the Kalman gain. The next step is to actually measure the process to obtain and then to generate and a posterior state estimate by incorporating the measurement as in equation (2.22). The final step is to obtain a posterior error covariance estimate (2.23). After each time and measurement update pair, the process is repeated with the previous posterior estimates used to project or predict the new a priori estimates. This recursive nature is one of the very appealing features of the Kalman filter. Figure (2.9) below offers a complete picture of the operation of the filter, combining figure (2.10)

2.6.6 Filter Parameter And Tuning

In the actual implementation of the filter, the measurement noise covariance is usually measured prior to the operation of the filter. Measuring the measurement error covariance is generally practical (possible) because it is possible to measure the process anyway (while operating the filter) so generally it is possible to take

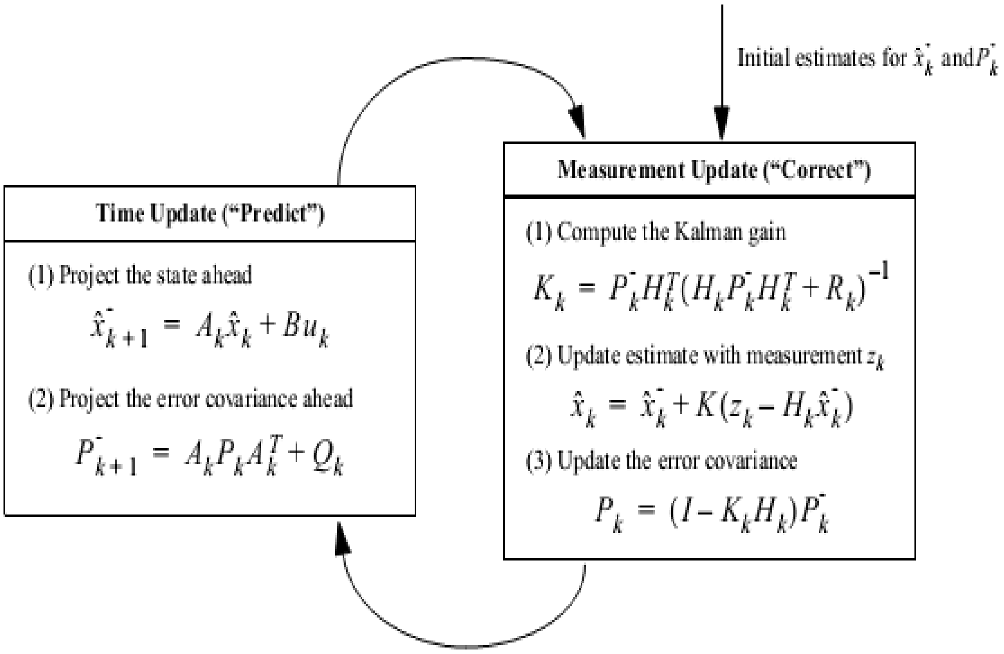
Initial estimates for $\hat{x}_k^-$ and P_k^-
Time Update ("Predict")
(1) Project the state ahead
$\hat{x}_{k+1}^- = A_k\hat{x}_k + Bu_k$
(2) Project the error covariance ahead
$P_{k+1}^- = A_kP_kA_k^T + Q_k$
Measurement Update ("Correct")
(1) Compute the Kalman gain
$K_k = P_k^-H_k^T(H_kP_k^-H_k^T + R_k)^{-1}$
(2) Update estimate with measurement z_k
$\hat{x}_k = \hat{x}_k^- + K(z_k - H_k\hat{x}_k^-)$
(3) Update the error covariance
$P_k = (I - K_kH_k)P_k^-$

Figure 2.8: Operation Of kalman Filter cycle[34]

some off-line sample measurements in order to determine the variance of the measurement noise. The determination of the process noise covariance is generally more difficult as it not feasible to directly observe the process estimation. Sometimes a relatively simple process model can produce acceptable results if one "injects" enough uncertainty into the process via the selection of. Certainly in this case one would hope that the process measurements are more reliable. It is frequently the case that the measurement error is not constant and the process noise is also sometimes changes during filter operation. The main problem with Kalman filtering is that statistical models are required for the system and the measurement instruments. Unfortunately, they are typically not available, or difficult to obtain. The two most commonly recommended methods of approaching this problem are:

- Employ an adaptive algorithm which adjusts these unknown parameters (such as the measurement noise variance) after each time step based on the observed measurements. This also accounts for processes with changing parameters.
- Perform an online analysis of the system and measurement instruments prior to running the process (system identification). It should be noted however that the second approach will not always be applicable if the process cannot be observed directly. In other words, if the measurements in the online analysis also contain errors, the process cannot be accurately profiled.

2.7 Motion model

Motion models are a simple type of predictor and are quite common among simple systems. Motion models aims is to predict the next position based on a number of past observations. There may or may not make use of acceleration and can be expressed as

$$P\left(t+1\right) = P\left(t\right) + V\left(t\right) \tag{2.24}$$

where $P\left(t+1\right)$ is the expected position at the next time step, $P\left(t\right)$ is the position at current time step, and $V\left(t\right)$ is the velocity at the current time step. For the simplest implementation,

$$V\left(t\right) = P\left(t\right) - P\left(t-1\right) \tag{2.25}$$

Chapter 3

Literature Survey

3.1 Introduction

The research conducted so far for object detection and tracking objects in video surveillance system are disscussed in this chapter. The set of challenges outlined above span several domains of research and the majority of relevant work will be reviewed in the upcoming chapters. In this section, only the representative video surveillance systems are discussed for better understanding of the fundamental concept. Tracking is the process of object of interest within a sequence of frames, from its first appearance to its last. The type of object and its description within the system depends on the application. During the time that it is present in the scene it may be occluded by other objects of interest or fixed obstacles within the scene. A tracking system should be able to predict the position of any occluded objects.

Object tracking systems are typically geared towards surveillance application where it is desired to monitor people or vehicles moving about an area. There are two district approaches to the tracking problem, top-down and another one is bottom-up. Top-down methods are goal oriented and the bulk of tracking systems are designed in this manner. These typically involve some sort of segmentation to locate region of interest, from which objects and features can be extracted for the tracking system. Bottom-up respond to stimulus and have according to observed changes. The top-down approach is most popular method for developing surveillance system. System has a common structure consisting of a segmentation step, a detection step, and a tracking step.

3.2 Literature Survey

As per the description in Chapter 1, object tracking has a lot of application in the real world. But it has many technological lacuna still exist in the methods of background subtraction. In this section, some previous works is disscussed for frame difference that use of the pixel-wise differences between two frame images to extract the moving regions, Gaussian mixture model based on background model to detect the object and finally background subtraction to detect moving regions in an image by taking the difference between current and reference background image in a pixel-by-pixel, and previous works done for the background modeling.

After the detection scenario is over, tracking part is done. Once the interesting objects have been detected it is useful to have a record of their movement over time. So tracking can be defined as the problem of estimating the trajectory of an object as the object moves around a scene. It is necessary to know where the object is in the image at each instant in time. If the objects are continuous observable and their sizes or motion does not vary over time, then tracking is not a hard problem.

In general surveillance systems are required to observe large area like airports, shopping malls. In these scenarios, it is not possible for a single camera to observe the complete area of interest because sensor resolution is finite and structures in the scene limit the visible area. Therefore surveillance of wide areas requires a system with the ability to track objects while observing them through multiple cameras. But here no disscussion about multiple camera network is done.

Lipton et al. [5] proposed frame difference that use of the pixel-wise differences between two frame images to extract the moving regions. In another work, Stauffer & Grimson et al. [6] proposed a Gaussian mixture model based on background model to detect the object. Liu et al. [7] ,proposed background subtraction to detect moving regions in an image by taking the difference between current and reference background image in a pixel-by-pixel. Collins et al. [8], developed a hybrid method that combines three-frame differencing with an adaptive background subtraction model for their VSAM (Video Surveillance and Monitoring) project. Desa & Salih et al [9], proposed a combination of background subtraction and frame difference that improved the previous results of background subtraction and frame difference. Sugandi et al. [10], proposed a new technique for object detection employing frame difference on low resolution image. Julio cezar et al. [3] has proposed a background model, and incorporate a novel technique for shadow detection in gray scale video sequences. Satoh et al. [11], proposed a new technique for object tracking employing block matching algorithm based on PISC image. Sugandi et al. [12], proposed tracking technique of moving persons using camera peripheral increment sign correlation image. Beymer & konolige et al. [2],1999 proposed in stereo camera based object tracking, use kalman filter for predicting the objects position and speed in x-2 dimension. Rosals & sclaroff et al.,1999 proposed use of extended kalman filter to estimate 3D trajectory of an object from 2D motion.

In object detection method, many researchers have developed their methods. Liu et al., 2001 proposed background subtraction to detect moving regions inan image by taking the difference between current and reference background image in a pixel-by-pixel. It is extremely sensitive to change in dynamic scenes derived from lighting and extraneous events etc. In another work, Stauffer & Grimson, 1997 proposed a Gaussian mixture model based on background model to detect the object. Lipton et al., 1998 proposed frame difference that use of the pixel-wise differences between two frame images to extract the moving regions. This method is very adaptive to dynamic environments, but generally does a poor job of extracting all the relevant pixels, e.g., there may be holes left inside moving entities. In order to overcome disadvantage of two-frames differencing, in some cases three-frames differencing is used. For instance, Collins et al., 2000 developed a hybrid method that combines three-frame differencing with an adaptive background subtraction model for their VSAM (Video Surveillance and Monitoring) project. The hybrid algorithm successfully segments moving regions in video without the defects of temporal differencing and background subtraction. Desa & Salih, 2004 proposed a combination of background subtraction and frame difference that improved the previous results of background subtraction and frame difference.

In object tracking methodology, this article will describe more about the region based tracking. Region-based tracking algorithms track objects according to variations of the image regions corresponding to the moving objects. For these algorithms, the background image is maintained dynamically and motion regions are usually detected by subtracting the background from the current image. Wren et al., 1997 explored the use of small blob features to track a single human in an indoor environment. In their work, a human body is considered as a combination of some blobs respectively representing various body parts such as head, torso and the four limbs. The pixels belonging to the human body are assigned to the differen t body part's blobs. By tracking each small blob, the moving human is successfully tracked. McKenna et al., 2000 proposed an adaptive background subtraction method in which color and gradient information are combined to cope with shadows and unreliable color cues in motion segmentation. Tracking is then performed at three levels of abstraction: regions, people, and groups. Each region has a bounding box and regions can merge and split. A human is composed of one or more regions grouped together under the condition of geometric structure constraints on the human body, and a human group consists of one or more people grouped together.

Cheng & Chen, 2006 proposed a color and a spatial feature of the object to identify the track object. The spatial feature is extracted from the bounding box of the object. Meanwhile, the color features extracted is mean and standard value of each object. Czyz et al., 2007 proposed the color distribution of the object as observation model. The similarity of the objects measurement using Bhattacharya distance. The low Bhattacharya distance corresponds to the high similarity.

To overcome the related problem described above, this article proposed a new technique for object detection employing frame difference on low resolution image Sugandi et al., 2007, object tracking employing block matching algorithm based on PISC image Satoh et al., 2001 and object identification employing color and spatial information of the tracked object Cheng & Chen, 2006.

3.3 Simulative Result of Detection and Tracking Algorithm

This chapter gives the idea about the existing and some modified methods analysis of algorithms for detection and tracking of objects. First some existing algorithm for detecting the objects like Frame difference method, Gaussian Mixture model to detect the object is disccused. Finally, background substrction and background modelling is shown. After implementing all these exiesting algorithms then put one modified model for background modelling.

Tracking is the process of object of interest within a sequence of frames,from its first appearance to its last.The type of object and its description within the system depends on the application.During the time that it is present in the scene it may be occluded by other objects of interest or fixed obstacles within the scene.A tracking system should be able to predict the position of any occluded objects.

Object tracking systems are typically geared towards survillance application where it is desired to monitor people or vehicles moving about an area.The ball tracking system has become a stadard feature of tenise and cricket broadcast and uses object tracking techniques to locate and track the ball as it moves in the court.

First implementation of an existing algorithm for tracking the object by using Block matching method is done.

3.4 Motion Detection

An automatic video surveillance is used by private companies, governments and public organizations to fight against terrorism and crime, public safety in airports, bus stand, railway station, town centers and hospitals. It has also find applications in traffic surveillance for efficient management of transport networks and road safety. Video surveillance system include task such as motion detection,tracking, and activity recognition. Out of the task mentioned above, detection of moving object is the first important step and successful segmentation of moving foreground object from the background ensures object classification, personal identification, tracking, and activity analysis, making these later step more efficient. Hu et al. [13] categorized motion detection into three major classes of method as frame differencing,background subtraction and Gaussian mixture.

3.5 Frame difference method

Frame differencing is a pixel-wise differencing between two or three consecutive frames in an image sequence to detect regions corresponding to moving object such as human and vehicles. The threshold function determine's change and it depends on the speed of object motion. It's hard to maintain the quality of segmentation, if the speed of the object changes significantly. Frame differencing is very adaptive to dynamic environments, but very often holes are developed inside moving entities.

Videos are actually consists of sequences of images, each of which called as a frame. For detecting moving objects in video surveillance system, use of frame difference technique from the difference between the current frame and a reference frame called as 'background image' is shown. That method is known as frame difference method.

Frame differencing is the simplest moving object detection method which is based on determining the difference between input frame intensities and background model by using pixel per pixel subtraction.

Grad. Sch. of Eng. et al. [5] have proposed frame difference method to detect the moving objects. In this case, frame difference method is performed on the three successive frames, which are between frame F_k and F_{k-1} and also the frame between F_k & F_{k+1} and the output image as frame difference image is two difference images d_{k-1} and d_{k+1} is expressed as

$$d_{k-1} = |f_k - f_{k-1}| \tag{3.1}$$

$$d_{k+1} = |f_k - f_{k+1}| \tag{3.2}$$

$$d_{k'}(x,y) = \begin{cases} 1, \; if \; d_{k'}(x,y) > T \\ 0, \; otherwise \end{cases} \tag{3.3}$$

Where $k' \; = \; k-1 \; and \; k+1$

The process is followed by applying and operator to d_{k-1} and d_{k-1} This method is already discussed in details in chapter (2). Here original frames are shown and after preprocessing segmented results frames are also shown.

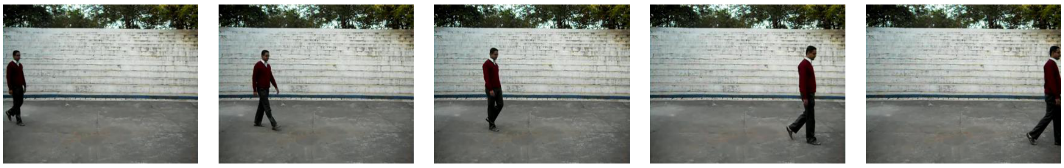

Figure 3.1: Original video frames

3.6 Background subtraction

The background subtraction [10] is the most popular and common approach for motion detection. The idea is to subtract the current image from a reference

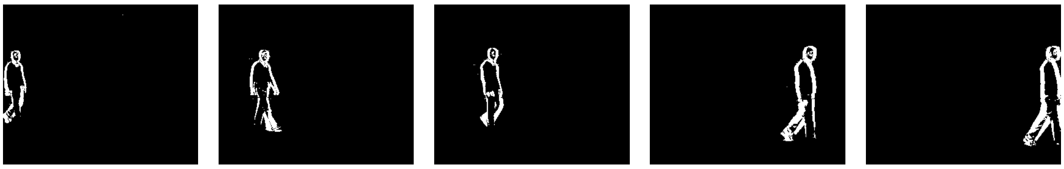

Figure 3.2: Output after frame difference

background image, which is updated during a period of time. It works well only in the presence of stationary cameras. The subtraction leaves only non-stationary or new objects, which include entire silhouette region of an object. This approach is simple and computationally affordable for real-time systems, but are extremely sensitive to dynamic scene changes from lightning and extraneous event etc. Therefore it is highly dependent on a good background maintenance model.

Here in this chapter simulation of different background subtraction techniques available in the literature, for motion segmentation of object is performed. Background subtraction detects moving regions in an image by taking the difference between the current image and the reference background image captured from a static background during a period of time. The subtraction leaves only non-stationary or new objects, which include entire silhouette region of an object. The problem with background subtrac-tion [14], [8] is to automatically update the background from the incoming video frame and it should be able to overcome the following problems:

- **Motion in the background**: Non-stationary background regions, such as branches and leaves of trees, a flag waving in the wind, or flowing water, should be identified as part of the background.

- **Illumination changes:** The background model should be able to adapt, to gradual changes in illumination over a period of time.

- **Memory**: The background module should not use much resource, in terms of computing power and memory.

- **Shadows**: Shadows cast by moving object should be identified as part of the background and not foreground.
- **Camouflage**: Moving object should be detected even if pixel characteristics are similar to those of the background
- **Bootstrapping**: The backgroundmodel should be able to maintain background even in the absence of training background (absence of foreground object).

3.7 Simple Background Subtraction

In simple background subtraction a absolute difference is taken between every current image $I_t(x, y)$ and the reference background image $B(x, y)$ to find out the motion detection mask $D(x, y)$. The reference background image is generally the first frame of a video, without containing foreground object.

$$D(x,y) = \begin{cases} 1, & if|I_t(x,y) - B(x,y)| \geq \tau \\ 0, & otherwise \end{cases} \tag{3.4}$$

where τ is a threshold, which decides whether the pixel is foreground or background. If the absolute difference is greater than or equal to τ , the pixel is classified as foreground, otherwise the pixel is classified as background.

3.8 Running Average

Simple background subtraction cannot handle illumination variation and results in noise in the motion detection mask. The problem of noise can be overcome, if the background is made adaptive to temporal changes and updated in every frame.

$$B_t(x,y) = (1-\alpha)B_{t-1}(x,y) + \alpha I_t(x,y) \tag{3.5}$$

where α is a learning rate. The binary motion detection mask D(x, y) is calculated as follows

$$D(x,y) = \begin{cases} 1, & if|I_t(x,y) - B(x,y)| \geq \tau \\ 0, & otherwise \end{cases} \tag{3.6}$$

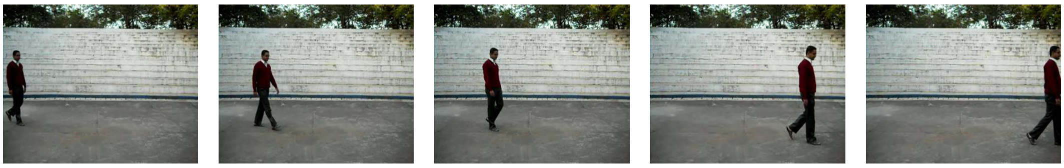

Figure 3.3: Original video frames

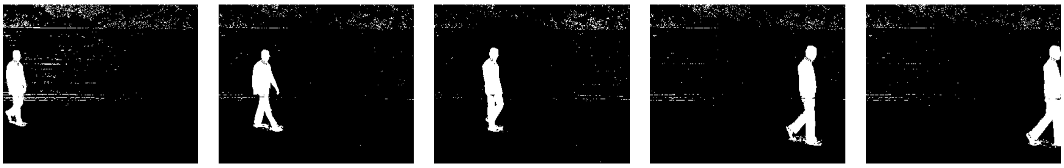

Figure 3.4: Output after background subtraction

3.9 Morphological Operation

Morphological operations apply a structuring element to an input image, creat-ing an output image of the same size. Morphological operation is performed to fill small gaps inside the moving object and to reduce the noise remained in the moving objects. The morphological operators implemented are dilation followed by erosion.In dilation, each background pixel that is touching an object pixel is changed into an object pixel. Dilation adds pixels to the boundary of the object and closes isolated background pixel. Dilation of set A by structuring element B [7] is defined as :

$$A \oplus B = \bigcup_{b \epsilon B} (A)_b \tag{3.7}$$

In erosion, each object pixel that is touching a background pixel is changed into a background pixel. Erosion removes isolated foreground pixels. Erosion of set A by structuring element B [7] is defined as:

$$A \ominus B = \bigcup_{b \epsilon B} (A)_{-b} \tag{3.8}$$

The number of pixels added or removed from the objects in an image depends on the size and shape of the structuring element used to process the image. Morphological operation eliminates background noise and fills small gaps inside an object. There is no fixed limit on the number of times dilation and erosion is performed. In the given algorithm dilation and erosion is used iteratively till the foreground object is completely segmented from the background.After morphological operation now the results of following frames, remove noise from frame difference and background subtraction frame result.

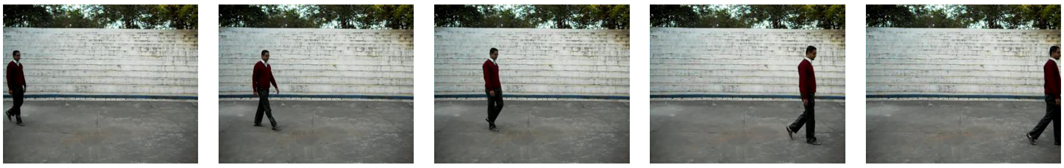

Figure 3.5: Original video frames

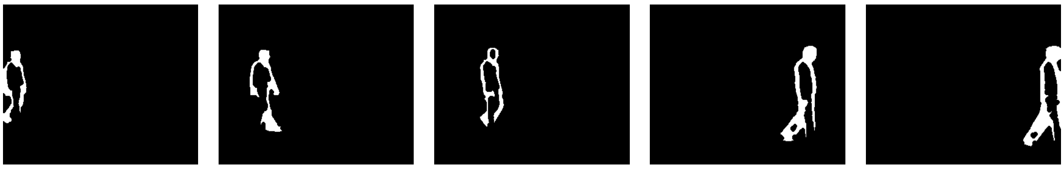

Figure 3.6: Output of frame difference reult after noise removal

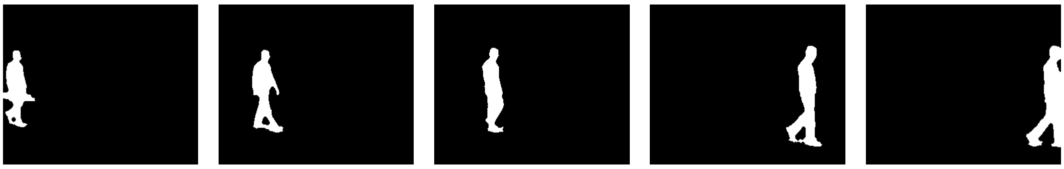

Figure 3.7: Output of background substraction reult after noise removal

3.10 Gaussian Mixture Model

To implement an existing Gaussian mixture model based on background model to detect the moving objects. For detecting moving objects in video surveillance system the use the Gaussian mixture model, is essential this model has the color values of a particular pixel as a mixture of Gaussians. But the pixel values that don't fit the background distributions are considered as foreground. Nowak 2003 showed how the parameters of a mixture of Gaussians for which each node of a sensor network had different mixing coefficients could be estimated using a distributed version of the well-known expectation-maximization (EM) algorithm. This message-passing algorithm involves the transmission of sufficient statistics between neighboring nodes in a specific order, and was experimentally shown to converge to the same results as centralized EM. Kowalczyk and Vlas-sis Kowalczyk and Vlassis, 2004 proposed a related gossip-based distributed algorithm called Newscast EM for estimating the parameters of a Gaussian mixture. Random pairs of nodes repeatedly exchange their parameter estimates and combine them by weighted averaging.

In this section, another technique that is commonly used for performing background segmentation. Stauffer and Grimson et al. [5]have proposed ,suggest a probabilistic approach using a mixture of Gaussian for identifying the background and foreground objects. The probability of observing a given pixel value P_t at time t is given by

$$P(p_t) = \sum_{i=1}^{k} \omega_{i,t}\eta(p_t, \mu_{i,t}, \sum i, t) \tag{3.9}$$

Where k is the number of Gaussian Mixture and that is used. The number of k varies depending on the memory allocated for simulations.Then the normalized Gaussian η is a function of

$\omega_{i,t}, \mu_{i,t}, \sum i, t$ which represents weight, mean and co-variance matrix of the ith Gaussian at time respectively.The weight indicates the influence of the ith Gaussian and time t. In this case k=5 to maximize the distinction amongst pixel values.Since it is an iterative process that all parameters are updated, with the inclusion of every new pixel. Before update take place, then the new pixel is compared to see if it matches any of the k existing Gaussian. A match is determined if $|p_t - \mu_{i,t}| < 2.5\sigma$

Where correspond to the standard deviation of the Gaussian. Depending on the match, the Gaussian mixture is updated in the following manner:

$$\omega_{i,t} = (1-\alpha)\omega_{i,t-1} + \alpha \tag{3.10}$$

$$\mu_{i,t} = (1-\rho)\mu_{t-1} + p_t\,\rho \tag{3.11}$$

$$\sigma^2_{i,t} = (1-\rho)\sigma^2_{i,t-1} + \rho(p_t - \mu_t)^T(p_t - \mu_t) \tag{3.12}$$

where

$$\rho = \alpha\eta(p_t|\mu_{i,t-1}, \sigma_{i,t-1}) \tag{3.13}$$

In this case the variable $(1|\alpha)$ defines the speed at which the distribution parameter changes. In the pixel (p_t) matches the i-th Gaussian, then the matching remaining (k-1) Gaussians are updated in the following manner,

$$\omega_{i,t} = (1-\alpha)\omega_{i,t-1} \tag{3.14}$$

$$\mu_{i,t} = \mu_{i,t-1} \tag{3.15}$$

$$\sigma^2_{i,t} = \sigma^2_{i,t-1} \tag{3.16}$$

The values for weight and variance vary based on the significance that is given to a pixel which is least likely to occure in a perticular way.

All the Gaussian weights are normalized after the update is performed.The k-Gaussians are then reordered based on their likelihood of exiestence.

Then (b) distribution are modeled to be the background and the remaining (k-b) distributionsa are modeled as the foreground for the next pixel.

The values for (b) is determined

$$B = argmin_b \left(\sum_{i=1}^{b} \omega_i > T \right) \tag{3.17}$$

Where T is some threshold value which measures the propotion of the data that needs to match the background and then the first B distribution are choosen as background model.

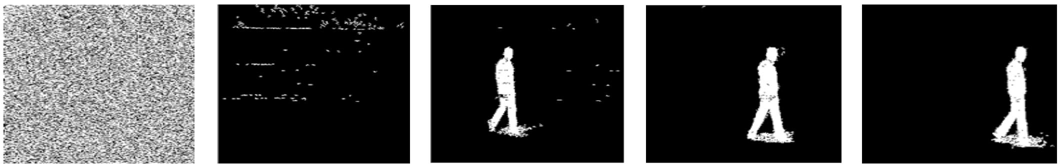

Figure 3.8: Gaussian mixture

3.11 W^4 background subtraction

W^4 model is a simple and effective method for segmentation of foreground ob-jects from video frame. In the training period each pixel uses three values; minimum m(x,y), maximum n(x,y) and the maximum intensity difference of pixels in the con-secutive frames d(x,y) for modeling of the background scene.The initial background for a pixel location (x, y) is given by [15] $\begin{bmatrix} m(x,y) \\ n(x,y) \\ d(x,y) \end{bmatrix} =$

$$\begin{bmatrix} min_z V^z(x,y) \\ max_z V^z(x,y) \\ max_z |V^z(x,y) - V^{z-1}(x,y)| \end{bmatrix}$$

Where 'z' are the frames satisfying

$|V^z(x,y)\lambda(x,y)| \leq 2\sigma(x,y)$

The background cannot remain same for a long period of time, so the initial back-ground needs to be updated. W^4 uses pixel-based update and object-based update method to cope with illumination variation and physical deposition of object. W^4 uses change map for background updation.A detection support map (gS) computes the number of times the pixel (x, y) is classified as background pixel.A detection support map g^S computes the number of times the pixel (x, y) is classified as background pixel.

$$gS_t(x,y) = \begin{cases} gS_{t-1}(x,y) + 1 \quad if\ pixel\ is\ background; \\ \\ \quad gS_{t-1}(x,y) \\ \qquad\qquad\qquad\qquad if\ pixel\ is\ foreground; \end{cases} \tag{3.18}$$

The background cannot remain same for a long period of time, so the initial back-ground needs to be updated.W^4 uses pixel-based update and object-based update method to cope with illumination variation and physical deposition of object.W^4 uses change map for background updation.

A detection support map (gS) computes the number of times the pixel (x, y) is classified as background pixel.

$$gS_t(x,y) = \begin{cases} gS_{t-1}(x,y) + 1 \quad if\ pixel\ is\ background \\ \\ \quad gS_{t-1}(x,y) \qquad if\ pixel\ is\ foreground \end{cases} \tag{3.19}$$

A motion support map (mS) computes the number of times the pixel(x, y) is classified as moving pixel.

$$mS_t(x,y) = \begin{cases} mS_{t-1}(x,y) + 1 & if\ M_t(x,y) = 1; \\ \quad mS_{t-1}(x,y) & if\ M_t(x,y) = 0; \end{cases} \tag{3.20}$$

where

$$M_t(x,y) = \begin{cases} 1\ if(|I_t(x,y) - I_{t+1}(x,y)| > 2 * \sigma) \wedge \\ \qquad (|I_{t-1}(x,y) - I_t(x,y) > 2 * \sigma) \\ \qquad\qquad 0\ \ otherwise \end{cases} \tag{3.21}$$

3.12 Background Modeling

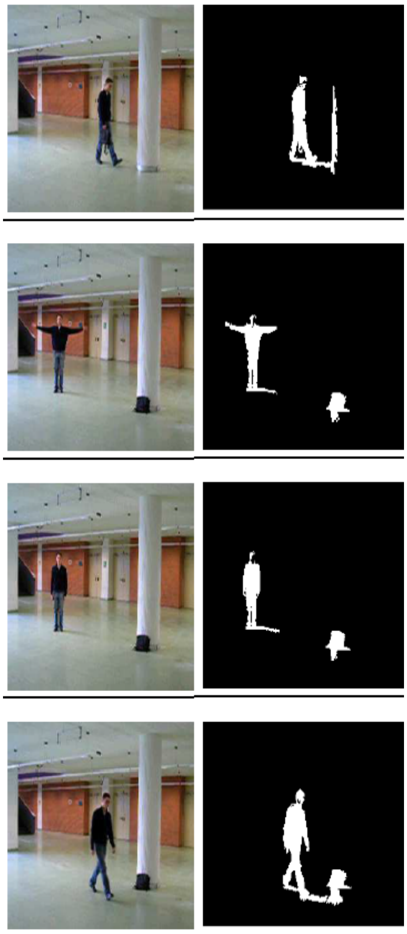

Figure 3.9: Output after background substraction using Julio cezar method

The basic principle of background subtraction is to compare a static background frame with the current frame of the video scene pixel by pixel. This technique builds a model of the background any frame can be compared with the model to detect zones where a significant difference occurs. Some steps for the background subtraction, i.e. first developement of a background model of the scene is done, then background subtraction to detect foreground object.

Julio cezar et al. [3] has proposed a background model, and incorporates a novel technique for shadow detection in gray scale video sequences. In the first stage, a pixel wise median filter over time is applied to several seconds of videos (typically

20-40 seconds) to distinguish moving pixels from stationary pixels.

In the second stage, only those stationary pixels are processed to construct the initial background model. Let V be an array containing N consecutive images,$V^k(i,j)$ be the intensity of a pixel (i,j) in the K-th images of $V, \sigma(i,j)$ and $\lambda(i,j)$ be the standard deviation and median value of intensities at pixel (i,j) in all images in V, respectively. The initial background for a pixel (i,j) is performed by a three-dimensional vector: the minimum m(i,j) and maximum n(i,j) intensity values and the maximum intensity difference d(i,j) between the consecutive frames observed during this training period. Then the background model B(i,j)=[m(i,j), n(i,j), d(i,j)] is obtained

As follows: $$\begin{bmatrix} m(i,j) \\ n(i,j) \\ d(i,j) \end{bmatrix} = \begin{bmatrix} min_z V^z(i,j) \\ max_z V^z(i,j) \\ max_z |V^z(i,j) - V^{z-1}(i,j)| \end{bmatrix}$$

Where 'z' are the frames satisfying $|V^z(i,j)\lambda(i,j)| \leq 2\sigma(i,j)$

After the training period, an initial background model B(i,j) is then obtained. Then each input image $I_t(i,j)$ of the video sequence is compared to B(i,j), and a pixel (i,j) is classified as a foreground pixel if $I^t(i,j) > (m(i,j) - k\mu)\, and\, I^t(i,j) < (n(i,j) + k\mu)$

After the background subtraction by using the above method, some shadow region pixels which is wrongly classified as foreground objects.so they have employed a shadow region detection method to remove the shadow pixels from detected foreground pixels.

3.12.1 Analysis

Here analysis of different detection algorithm is done. In the experiment of taking different video, it is shown that, Julio cezar method is the best object detection method. After getting the silhouette of the background objects, the contour of the foreground objects after some post processing of the resulting silhouette like region filling is performed. Then extraction of the feature from the contour is done.But the feature extraction part is not discussed here.

3.13 Object Tracking

After the object detection is achieved, the problem of establishing a correspondence between object masks in consecutive frames should arise. Obtaining the correct track information is crucial for subsequent actions, such as object identification and activity recognition. For this situation, block matching technique is used.

3.13.1 Block Matching Method

The entire process of tracking the moving object is illustrated in the following Fig 2.6. The block matching method is well described in [4], which we have applied here.

Block matching is a technique for tracking the interest moving object among the moving objects emerging in the scene. In this article, the blocks are defined by dividing the image frame into non-overlapping square parts. The blocks are made based on peripheral increment sign correlation (PISC) image Satoh et al., 2001; Sugandi et al., 2007 that considers the brightness change in all the pixelsof the blocks relative to the considered pixel. Fig. 9 shows the block in PISC image with block size is 55 pixels. Therefore, one block consists of 25 pixels. The blocks of the PISC image in the previous frame are defined as shown in Eq. (2.3). Similarly, the blocks of the PISC image in the current frame are defined in Eq. (2.4). To determine the matching criteria of the blocks in two successive frames, evaluation is done using correlation value that expresses in Eq. (2.5). This equation calculates the correlation value between block in the previous frame and the current one for all pixels in the block. The high correlation value shows that the blocks are matched each other. The interest moving object is determined when the number of matching blocks in the previous and current frame are higher than the certain threshold value. The threshold value is obtained experimentally.

$$b_{np} = \begin{cases} 1, & iff_{np \geq f(i,j)} \\ o, & otherwise \end{cases} \tag{3.22}$$

$$b'_{np} = \begin{cases} 1, & iff_{np \geq f(i,j)} \\ o, & otherwise \end{cases} \tag{3.23}$$

$$corr^{n} = \sum_{P=0}^{N} b_{np} * b'_{np} + \sum_{P=0}^{N} (1 - b_{np}) * (1 - b'_{np}) \tag{3.24}$$

where p & p' are the block in the previous and current frame, n is the number of block and N is the number of pixels of block.

3.13.2 Tracking Method

The tracking method used in this article can be described as following. The matching process is illustrated in Fig.2.6. Firstly, blocks and the tracking area are made only in the area of moving object to reduce the processing time. The previous frame is devided into block size (block A) with 9x9 pixels in the previous frame. It is assume that the object coming firstly will be tracked as the interest moving object. The block A will search the matching block in each block of the current frame by using correlation value as expresses in Eq.(2.5). In the current frame, the interest moving object is tracked when the object has maximum number of matching blocks. When that matching criteria is not satisfied, the matching process is repeated by enlarging the tracking area (the rectangle with dash line).The blocks still are made inside the area of moving object. When the interest moving object still cannot be tracked, then the moving object is categorized as not interest moving object or another object and the tracking process is begun again from the begin.

3.13.3 Feature Extraction

The feature of objects extracted inthe spatial domain is the position of the tracked object. The spatial information combined with the features in time domain represents the trajectory of the tracked object, so the movement and speed of the moving objects can be estimated that needs to tracked. Therefore, the features of spatial domain are very important toobject identification. The bounding box defined in Eq. (2.4) is used as spatial information of moving objects.

After getting the interest moving object,then extraction of interest moving object by using a bounding box. The bounding box can be determined by computing the maximum and minimum value of x and y coordinates of the interest moving object according to the following equation:

$$B^i_{min} = \{(x^i_{min}, y^i_{min})|x, y \in O^i\} \tag{3.25}$$

$$B^i_{max} = \{(x^i_{max}, y^i_{max})|x, y \in O^i\} \tag{3.26}$$

where O^i denotes set of coordinate of points in the interest moving object i,B^i_{min} is the left top corner cordinates of the interest moving object i, and B^i_{max} is the right bottom corner cordinates of the interesting moving object i.In the chapter 4 shows the bounding box of the object tracking.

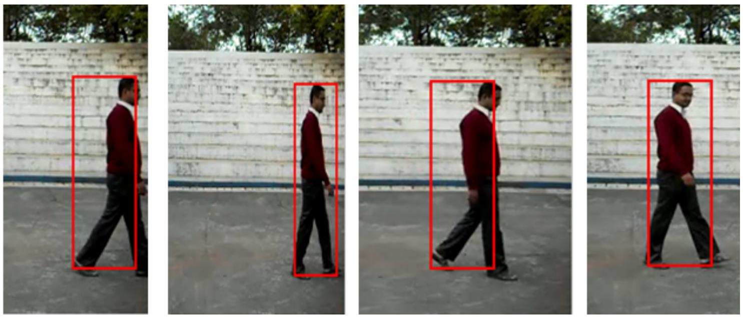

Figure 3.10: Video tracking results

3.14 Motion-Based Multiple Object Tracking

Detection of moving objects and motion-based tracking are important components of many computer vision applications, including activity recognition, traffic monitoring, and automotive safety. The problem of motion-based object tracking can be divided into two parts

- Detecting moving objects in each frame
- Associating the detections corresponding to the same object over time

The detection of moving objects uses a background subtraction algorithm based on Gaussian mixture models. Morphological operations are applied to the resulting foreground mask to eliminate noise. Finally, blob analysis detects groups of connected pixels, which are likely to correspond to moving objects.

The association of detections to the same object is based solely on motion. The motion of each track is estimated by a Kalman filter. The filter is used to predict the track's location in each frame, and determine the likelihood of each detection being assigned to each track.

In any given frame, some detections may be assigned to tracks, while other detections and tracks may remain unassigned.The assigned tracks are updated using the corresponding detections. The unassigned tracks are marked invisible. An unassigned detection begins a new track.

In motion detection how multiple objects are tracked by using Kalman filter.At first system objects are created for reading the video frames.Then kalman filter objects are used for motion based tracking.Then the total number of frames in which tracks are detected.Here it is shown, how motion based multiple objecs are tracked. The algorithm involves two steps:

- Step 1: Compute the cost of assigning every detection to each track using the distance method. The cost takes into account the Euclidean distance between the predicted centroid of the track and the centroid of the detection. It also includes the confidence of the prediction, which is maintained by the Kalman filter

- Step 2: Solve the assignment problem represented by the cost matrix using the assign Detections To Tracks function. The function takes the cost matrix and the cost of not assigning any detections to a track.

The value for the cost of not assigning a detection to a track depends on the range of values returned by the distance method of the KalmanFilter. This value must be tuned experimentally. Setting it too low increases the likelihood of creating a new track, and may result in track fragmentation. Setting it too high may result in a single track corresponding to a series of separate moving objects.

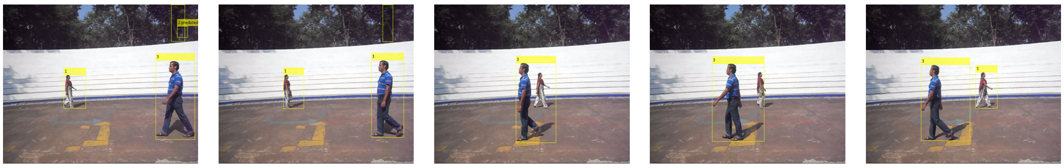

Figure 3.11: Original video frames

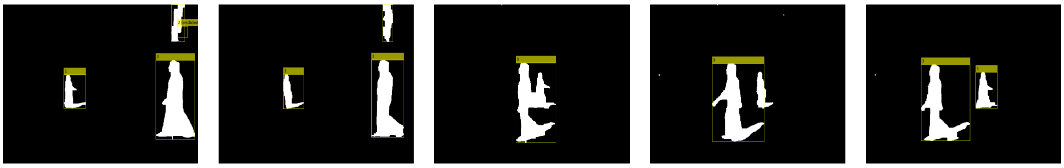

Figure 3.12: Multiple object tracking

3.15 Simulator

Matlab is a simple an event driven simulation tool which provides a platform to analyze the static and dynamic nature of the video processing. All experiments relevant to the thesis are carried out on 2.81GHz AMD Athlon 64 X2 Dual Core processor with 2GB RAM. The experiments are simulated using Matlab different Version 7.10.0.499 (R2012).In this chapter, the simulator and the simulation parameter that are used for experiments are discussed .

3.16 Conclusion

This chapter reviews the literature surveys that have been done during the re-search work. The related work that has been proposed by many researchers has been discussed . The research papers related to object detection and tracking of objects diagnosis from 1998 to 2011 has been shown which discussed about different methods and algorithm to diagnose the tracking system.

Chapter 4

Conclusions

4.1 Conclusions

In every chapter the object detection and tracking methods are being surveyed. This thesis has examined methods to improve the performance of motion segmentation algorithms and Block matching technique for object tracking applications and examined methods for multi-modal fusion in an object tracking system.

Motion segmentation is a key step in many tracking algorithms as it forms the basis of object detection. Improving segmentation results as well as being able to extract additional information such as frame difference, Gaussian of mixture model, background subtraction allows for improved object detection and thus tracking. However a strength of kalman filter is their ability to track object in adverse situation. Integrating a kalman filter within a standard tracking system allows the kalman filter is to use progressively updated features and aids in main training identity of the tracked object, and provides tracking system with an effective means. The simulator and the simulation parameters used for the experiments are disscussed. We have shown the simulation results in the form of images.

Bibliography

[1] Alper Yilmaz, Omar Javed, and Mubarak Shah. Object tracking: A survey. *Acm Computing Surveys (CSUR)*, 38(4):13, 2006.

[2] Gary Bishop and Greg Welch. An introduction to the kalman filter. *Proc of SIGGRAPH, Course*, 8:27599–3175, 2001.

[3] J Cezar Silveira Jacques, Claudio Rosito Jung, and Soraia 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.

[4] Budi Sugandi, Hyoungseop Kim, Joo Kooi Tan, and Seiji Ishikawa. A block matching technique for object tracking employing peripheral increment sign correlation image. In *Computer and Communication Engineering, 2008. ICCCE 2008. International Conference on*, pages 113–117. IEEE, 2008.

[5] 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.

[6] 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.

[7] Ya Liu, Haizhou Ai, and Guang-you Xu. Moving object detection and tracking based on background subtraction. In *Multispectral Image Processing and Pattern Recognition*, pages 62–66. International Society for Optics and Photonics, 2001.

[8] Changick Kim and Jenq-Neng Hwang. Fast and automatic video object segmentation and tracking for content-based applications. *Circuits and Systems for Video Technology, IEEE Transactions on*, 12(2):122–129, 2002.

[9] Shahbe Mat Desa and Qussay A Salih. Image subtraction for real time moving object extraction. In *Computer Graphics, Imaging and Visualization, 2004. CGIV 2004. Proceedings. International Conference on*, pages 41–45. IEEE, 2004.

[10] Budi Sugandi, Hyoungseop Kim, Joo Kooi Tan, and Seiji Ishikawa. Tracking of moving objects by using a low resolution image. In *Innovative Computing, Information and Control, 2007. ICICIC'07. Second International Conference on*, pages 408–408. IEEE, 2007.

[11] YUTAKA Sato, S Kaneko, and SATORU Igarashi. Robust object detection and segmentation by peripheral increment sign correlation image. *Trans. of the IEICE*, 84(12):2585–2594, 2001.

[12] Mahbub Murshed1/2, Md Hasanul Kabir1/2, and Oksam Chae1/2. Moving object tracking-an edge segment based approach. 2011.

[13] Weiming Hu, Tieniu Tan, Liang Wang, and Steve Maybank. A survey on visual surveillance of object motion and behaviors. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 34(3):334–352, 2004.

[14] Zhan Chaohui, Duan Xiaohui, Xu Shuoyu, Song Zheng, and Luo Min. An improved moving object detection algorithm based on frame difference and edge detection. In *Image and Graphics, 2007. ICIG 2007. Fourth International Conference on*, pages 519–523. IEEE, 2007.

[15] Ismail Haritaoglu, David Harwood, and Larry S. Davis. W¡ sup¿ 4¡/sup¿: real-time surveillance of people and their activities. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):809–830, 2000.

[16] Deep J Shah, Deborah Estrin, and Afrouz Azari. Motion based bird sensing using frame differencing and gaussian mixture. *Undergraduate Research Journal*, page 47, 2008.

[17] Budi Sugandi, Hyoungseop Kim, Joo Kooi Tan, and Seiji Ishikawa. Tracking of moving objects by using a low resolution image. In *Innovative Computing, Information and Control, 2007. ICICIC'07. Second International Conference on*, pages 408–408. IEEE, 2007.

[18] Intan Kartika and Shahrizat Shaik Mohamed. Frame differencing with post-processing techniques for moving object detection in outdoor environment. In *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on*, pages 172–176. IEEE, 2011.

[19] Robert T Collins, Alan Lipton, Takeo Kanade, Hironobu Fujiyoshi, David Duggins, Yanghai Tsin, David Tolliver, Nobuyoshi Enomoto, Osamu Hasegawa, Peter Burt, et al. *A system for video surveillance and monitoring*, volume 102. Carnegie Mellon University, the Robotics Institute Pittsburg, 2000.

[20] 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.

[21] Pakorn KaewTraKulPong and Richard Bowden. An improved adaptive background mixture model for real-time tracking with shadow detection. In *Video-Based Surveillance Systems*, pages 135–144. Springer, 2002.

[22] Cláudio Rosito Jung. Efficient background subtraction and shadow removal for monochromatic video sequences. *Multimedia, IEEE Transactions on*, 11(3):571–577, 2009.

[23] Yung-Gi Wu and Chung-Ying Tsai. The improvement of the background subtraction and shadow detection in grayscale video sequences. In *Machine Vision and Image Processing Conference, 2007. IMVIP 2007. International*, pages 206–206. IEEE, 2007.

[24] Muhammad Shoaib, Ralf Dragon, and Jorn Ostermann. Shadow detection for moving humans using gradient-based background subtraction. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 773–776. IEEE, 2009.

[25] Jianhua Ye, Tao Gao, and Jun Zhang. Moving object detection with background subtraction and shadow removal. In *Fuzzy Systems and Knowledge Discovery (FSKD), 2012 9th International Conference on*, pages 1859–1863. IEEE, 2012.

[26] T Thongkamwitoon, S Aramvith, and TH Chalidabhongse. An adaptive real-time background subtraction and moving shadows detection. In *Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on*, volume 2, pages 1459–1462. IEEE, 2004.

[27] Pakorn KaewTraKulPong and Richard Bowden. An improved adaptive background mixture model for real-time tracking with shadow detection. In *Video-Based Surveillance Systems*, pages 135–144. Springer, 2002.

[28] John Brandon Laflen, Christopher R Greco, Glen W Brooksby, and Eamon B Barrett. Objective performance evaluation of a moving object super-resolution system. In *Applied Imagery Pattern Recognition Workshop (AIPRW), 2009 IEEE*, pages 1–8. IEEE, 2009.

[29] Lloyd L Coulter, Douglas A Stow, Yu Hsin Tsai, Christopher M Chavis, Richard W McCreight, Christopher D Lippitt, and Grant W Fraley. A new paradigm for persistent wide area surveillance. In *Homeland Security (HST), 2012 IEEE Conference on Technologies for*, pages 51–60. IEEE, 2012.

[30] Shalini Agarwal and Shaili Mishra. A study of multiple human tracking for visual surveillance. *International Journal*, 5, 1963.

[31] Fatih Porikli. Achieving real-time object detection and tracking under extreme conditions. *Journal of Real-Time Image Processing*, 1(1):33–40, 2006.

[32] Huchuan Lu, Ruijuan Zhang, and Yen-Wei Chen. Head detection and tracking by mean-shift and kalman filter. In *Innovative Computing Information and Control, 2008. ICICIC'08. 3rd International Conference on*, pages 357–357. IEEE, 2008.

[33] Greg Welch and Gary Bishop. An introduction to the kalman filter, 1995.

[34] Bastian Leibe, Konrad Schindler, Nico Cornelis, and Luc Van Gool. Coupled object detection and tracking from static cameras and moving vehicles. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(10):1683–1698, 2008.

[35] Bastian Leibe, Konrad Schindler, Nico Cornelis, and Luc Van Gool. Coupled object detection and tracking from static cameras and moving vehicles. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(10):1683–1698, 2008.