CHAPTER 1

Introduction

1.1 Introduction

Visual Monitoring System is the most research topic in Today’s era. Visual observation in computer vision helps to analyze object’s activities easily. Today’s era of computer vision will completely remove traditional human operated Video Surveillance System. A major part of smart video surveillance system is characterized by perception, the robustness of a Smart Video Surveillance System (SVSS) is not only to sense the environment but also to interpret and act intelligently. Advancement in perception will lead to applications for defense and automated driving assistance. Nowadays researchers are working on object detection, object tracking, crowd analysis, pedestrian and vehicle identification to improve the security at the public places.

T. Reeve [2] and Rajiv Shah [3] have beautifully surveyed the penetration and importance of the surveillance system in United Kingdom and United States. They have reviewed that the large amount of surveillance data monitoring was done by human operators over a longer time and yet it doesn’t yields vigilant monitoring. Modern researchers are putting more concentration on real time processing of visual observations because of the tremendous growth of computers and low cost high resolution cameras. Over more than a decade researchers focus their intention not only to object detection and object tracking for smart surveillance system but also on a real time processing multiple cameras and even on more recent developments, they focus on 3D object detection and tracking. For 3D object detection and tracking people are using different approaches for the depth calculation. Some of the methods are using stereovision for the depth calculation using the disparity while some are using Monocular clues for 3D detection. There is clear tradeoff among two approaches of computational v/s depth precision. Some of the current methods completely estimate 6D pose of the object.
Figure 1.1 shows the primitive operations of the video surveillance system. The automated surveillance system requires different components. The incoming video sequences may be with or without the background so, it is necessary to generate background with the help of the subsequent frames or by means of the frame analysis. Primitive operation preprocessing is required to remove the dataset noises and outliers. One of the important component of every surveillance system is to establish the background model which is appropriate for every video sequences. The robustness of background modeling gives accurate foreground detection. Finally, tracking can be achieved with the help of the tracker.

A typical visual surveillance system consists with the help of static Pan-Tilt-Zoom (PTZ) video cameras and it will transmit video sequences to a central surveillance room or stores in a surveillance monitoring server. Such video sequences are being observed carefully by
human operators. If monotonous monitoring activity of an operator might miss some important incidences at that particular time then such loss in lapses during monitoring activities becomes challenging.

**FIGURE 1.2 Example of an Indoor Surveillance System**

**FIGURE 1.3 Example of an Outdoor Surveillance System**

Any smart surveillance system must be able to perceive and identify the new scene because a human operator is not able to change algorithm parameters every time. The system can operate in different environmental conditions like sudden changes in illumination, clutter background, occlusions etc. with minimum error. Such system is called robust video surveillance system.

Single camera surveillance system is suffered by many issues like occlusions, different silhouette of still and moving object shadows, weather effects, dynamic background scene etc. Though the single camera is suffered with all such constraints, still it is preferable to some extent because of its limited capability and its affordable price. To handle all such constraints use of multi camera system is preferable instead of single camera for surveillance system. Even 3D detection and tracking can be achieved by means of the single or multiview monocular video sequences. The depth maps can be generated by means of the stereoscopic or monocular clues.
Such a multiple camera systems are also suffered with false negative because of information loss that takes place at the time of conversions from RGB video sequences to Gray sequences. When foreground and backgrounds have nearest gray levels, different dark colors are converted to gray levels which leads to decrease in detection and tracking accuracies.

1.2 Motivation

Situation alertness is the key to security. There are basically three kinds of work that a security analysts needs to “See, Check and Track”. They have to identify the people and vehicle in space, locate the people and what kind of activity they do in space. They also uses chronological context so as to understand the statistics taken from the above knowledge.

Enhancing and ensuring a fair level of security across multiple scales of time and space in public places such as airports, railway station and at other places becomes extremely multifarious challenge for smart video surveillance system and it also enhances situational consciousness. There are multiple security challenges like screening system, database system, biometric system and video surveillance system for object tracking and verifying identity and also to monitor activities respectively.

Today video surveillance system focusing on compression for the purpose of storing and transmitting performs as either analog or digital video recorders.

Locating, identifying and learning the object behavior in video sequence requires three main steps.
Motivation

- Detecting the objects – foregrounds that are in motion.
- Detected objects to be tracked in consecutive frames.
- Object behavior recognition.

Smart video surveillance system queries fast and robust algorithm for estimating background, motion segmentation, object tracking and scene analysis and it also assist operator for important scene events. Smart and intelligent video surveillance is the most researched topic for the last decade because more importance is given to security and military applications [4].

Moving object detection is the primitive operation for scene analysis. For accurate object tracking one must be able to detect foreground precisely in every frame under different conditions. Object should be detected as soon as it appears in the frame. In some cases where the exact background cannot be achieved in the earlier stages, the algorithm requires additional frames to create background or extract the background information under non-stationary background and in such cases, we may lose the tracking accuracy for those frames. The robustness of every tracking algorithm depends on the successful detection of foreground. Some of the popular object detection techniques are background subtraction, frame difference and optical flow.

Frame difference algorithm detect object by evaluating consecutive frames, it gives less computational time but it cannot adopt dynamic background. Background subtraction can be implemented using simply subtracting out background frames from every frame and it also is not able to handle dynamic backgrounds. Optical flow is an alternative approach to deal with the different background constraints but such a technique requires more computational time as compared to that of other approaches. In some surveillance situation both the camera and foreground are moving, optical flow is the best approach to handle such situations and also to handle dynamicity of the background. Instead of the traditional background estimation approaches, statistical background approaches such as Gaussian Mixture Model gives more robustness and can also handle different backgrounds.

![Figure 1.5 Simple Background Estimation Approach](image-url)
Robust and accurate tracking of non-rigid, complex, fast moving objects being a major challenges now a days. Object tracking is a vital task in surveillance system. It is also applicable in video editing, augmented reality, traffic monitoring and control, gesture and posture recognition etc. The challenging task is to track object in different conditions like illumination variations, dynamic background, complex object silhouette and occlusions. A robust tracking must predict the positions of the object whether it is being occluded or not and ensuring its position and makes sure that it will not lost it completely. Multiple views and cameras can handle such conditions. One of the major motivation for tracking is the ball tracking systems ‘Hawk – eye’ [133]. It uses object tracking techniques to track ball in cricket as well as in tennis sport [134] and proposed two different tracking approaches named, top-down and bottom-up. Bottom up approach is generally focused on typical application like data mining, gesture recognition and sports events. While top-down is generally used for the surveillance purposes. In the first world country, smart and real time surveillance system uses top-down approach. Top-down approach consists foreground segmentation, motion detection and object tracking.

FIGURE 1.6 Simple Object Tracking Approach

FIGURE 1.7 Low Frame Rate Video Sequences – Consecutive Frames

Figure 1.7 shows Low Frame Rate video sequences. It would be very challenging for every video surveillance system to segment the moving foregrounds and track the moving objects. It indicates that a video sequence is challenging and gives ample amount of motivation to every researchers.
Tracking algorithm should also be robust in terms of not only handle the constraints but also to operate on both the environment such as indoor and outdoor. Template matching, mean shift, motion estimation, kalman filtering, particle filtering, silhouette tracking are some of the popular tracking approaches available for object tracking.

1.3 Objective

This research work aims to develop robust background modeling for the foreground detection in both the indoor and outdoor environment. The main objective of the thesis is to develop an algorithm which can detect and track 2D and Monocular 3D information in visual surveillance system using probabilistic statistical approach. Also to ensure high level of security in public places using static (PTZ) camera, robust detection and tracking algorithm for video sequences easily adopt background changes.

The aim of this work is to improve the performance evaluation of the detection and tracking system under different challenges. The proposed system provides possible intrinsic and extrinsic improvements to the challenges as mentioned below.

Intrinsic Improvements: Motion segmentation

- To select model parameters appropriately using parameter optimization algorithm.
- To develop a robust foreground detection algorithm through adaptive thresholding for motion segmentation.

Extrinsic Improvements: Performance evaluation

- To improve Performance evaluation parameter using post processing technique.
- To remove Dataset noise by preprocessing technique.

Moreover, the proposed work also targets a few additional constraints as follows,

- Performance under different backgrounds.
- Handling varying lighting conditions such as bright, dark, low and high contrast.
- Invariance to camera perspective.
Introduction

- Handling process and data noise.
- Working on various resolutions and video sequences with different framer rates.
- Handling partial occlusion and it is able to detect and track near field, mid field and far field objects.
- Detecting and tracking objects with similar appearance, different height and with the different motion.
- Detecting and tracking 3D multiple objects in monocular video sequences.
- Handling crowded scenes.

1.4 Contribution

This thesis proposed a unifying approach for 2D and monocular 3D object detection and tracking. Proposed approach integrates different modules like Modified Gaussian Mixture Model, Adaptive thresholding for the motion segmentation and Kalman filtering for both the indoor and outdoor surveillance systems. Proposed approach is unique and simpler in reference to other state-of-the-art approaches. It is capable of handling different background dynamics and constraints. Robustness in terms of handling the different constraints is the major goal and achievement of proposed work.

This research work provides four-fold contribution.

(i) Intrinsic Improvement – Parameter Optimization algorithm

Intrinsic Improvements in the Gaussian Mixture Model which concern the modification made in Gaussian Model parameter initialization and parameter maintenance during execution at every new pixel or frame level and also at the foreground detection (motion segmentation) level. The appropriate selection of mixture parameter is indeed an impact on the performance of the overall surveillance system, as the same algorithm is applicable for both indoor and outdoor surveillance system. Literature survey shows that usually, the model parameters are predefined or initialized by some algorithms like k-means cluster algorithm, EM or MLE approach etc. In proposed algorithm, model parameters can be initialized by parameter optimization algorithm for the every video sequences. This Proposed algorithm is evaluated with the standard video datasets and compared with the other similar approaches and the significant improvements are obtained as outcome.
(ii) **Intrinsic Improvement – Foreground Detection (Motion Segmentation)**

Foreground detection plays a vital role in surveillance system. Background model is sensitive enough to segment every moving object. Literature survey provides different Intensity, Region, Texture, Edge or Motion based segmentation approaches. Static threshold provides poor foreground detection and may lead to increases either false positive or false negative. In proposed approach foreground detection is achieved by means of Adaptive Thresholding instead of static thresholding. The proposed approach is evaluated with the standard dataset and resultant foreground mask is compared with the ground truth and other similar approaches. The observation is that most of the false negatives generated by the traditional pixel and region based methods are removed by the intrinsic improvements.

(iii) **Extrinsic Improvement – Pre processing**

Extrinsic Improvements emphasizes purely on improving the performance of the model and hence it also tends to improve the results. The image and dataset noises are removed by using the preprocessing. The proposed algorithm used Adaptive Local Noise Reduction Filter as a pre-processing method to remove dataset noises.

(iv) **Extrinsic Improvements – Post processing**

Post processing is again an external tool to perform the evaluation. In proposed algorithm Morphological Closing (dilation followed by erosion) is being used as a post-processing method for the sake of reducing the noise and outliers in the datasets. As a result most of the false positives generated by the traditional approaches are removed.

So far as performance evaluation is concerned, the proposed method shows significant improvements in well known and widely used parameters such as similarity measures, objects detection and tracking accuracy, Recall and Precision as compared to other similar pixel based parametric, nonparametric and region based background subtraction counterparts.
1.5 Thesis Organization

The problem in indoor and outdoor surveillance is investigated in the subsequent chapters and that is able to detect and track objects in both the environments and handle dynamic backgrounds and partial occlusion.

The chapters are structured as follows.

Chapter 2: Literature Review
- Provides brief and general literature survey on background model estimation, various parameter initialization algorithms, various background modeling methods, motion segmentation approaches and detailed investigation of object tracking approaches.

Chapter 3: Background Modeling
- Proposed Background model and necessary derivations related to probabilistic model for both the indoor and outdoor environment.
- Proposed Parameters Initialization algorithm and maintenance techniques.
- Pre and post processing techniques for performance enhancement.
- Foreground detection using adaptive thresholding.
- 3D monocular object detection.

Chapter 4: Object Tracking: Kalman Filtering
- Object Tracking concept and various approaches.
- Kalman filter mathematical expression and discussion.
- 3D monocular object tracking.

Chapter 5: Results and Discussion
- Object detection and tracking analysis.
- Comparison of performance evaluation for the other similar approaches with standard datasets.
- 3D Monocular object detection and tracking.
- Comparison of performance evaluation for 3D monocular object detection and tracking with the other similar approaches using standard datasets.
- Comparison of 2D and 3D approaches for the monocular scenes.
- Performance evaluation of the proposed approach is evaluated using various other metrics.
Chapter 6: Conclusion and Future Work

- Provides summary of the proposed approach with the comparative results and provides certain suggestions on supplementary investigation on proposed subject and area.

**Thesis Organization**

| Video Sequence – Video/ Image Processing (Frame Analysis – Background Analysis) |
| (Chapter 3) Preprocessing |
| (Chapter 3) Background Model Modified Gaussian Mixture Model |
| (Chapter 3) Fore Ground Detection |
| (Chapter 3) Post Processing |
| (Chapter 4) Object Tracking – Kalman Filtering |
| (Chapter 5) Results Performance Evaluation and Comparative Analysis |

**Original Contribution**

**Proposed Object Detection and Tracking Algorithm**

- Adaptive Local Noise Reduction Filter
- Background Model Parameter Initialization Algorithm
- Adaptive Thresholding Technique
- Morphological closing (Dilation followed by erosion)

**FIGURE 1.8 Thesis Organization - Original Contribution made towards Thesis**