MOVING FOREGROUND OBJECT DETECTION AND TRACKING IN VISUAL SURVEILLANCE SYSTEM

A DISSERTATION
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
SIGNAL PROCESSING AND DIGITAL DESIGN

Submitted by

Swati Kumari
2K13/SPD/19

Under the supervision of

Mr. Rajesh Rohilla
(Associate Professor)

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

2013-2015

Certificate

This is to certify that the dissertation title ***"Moving Foreground Object Detection and Tracking in Visual Surveillance System"*** submitted by **Ms. Swati Kumari**, Roll. No. *2K13/SPD/19*, in partial fulfilment for the award of degree of Master of Technology in Signal Processing & Digital Design at **Delhi Technological University, Delhi**, is a bonafide record of student's own work carried out by her under my supervision and guidance in the academic session 2014-15. To the best of my belief and knowledge the matter embodied in dissertation has not been submitted for the award of any other degree or certificate in this or any other university or institute.

(Rajesh Rohilla)
Associate Professor/Dept. of ECE
Delhi Technological University

Acknowledgement

I am indebted to my thesis supervisor Mr Rajesh Rohilla, Associate Professor Department of Electronics and Communication, for his gracious encouragement and very valued constructive criticism that has driven me to carry out the project successfully.

I am greatly thankful to Prof. Prem R. Chadda, Head of Department (Electronics & Communication Engineering), entire faculty and staff of Electronics & Communication Engineering and friends for their continuous support, encouragement and inspiration in the execution of this "thesis" work.

I would like to express my heartiest love and respect to my husband, Mr. Abhishek Srivastava and my family members, due to their blessing and moral support, I have completed my dissertation work.

Finally I express my deep sense of gratitude to my parents who bestowed upon me their grace and were source of my inspiration and encouragement.

Swati Kumari
2K13/SPD/19
M. Tech (SPDD)

Dedicated to My Parents

ABSTRACT

Moving object tracking has been an interesting research field since decades due to its wide applications in the field of security, surveillance, medicine, entertainment etc. A lot of research work has been done previously for object tracking when the camera remains stationary. But, now a days due to increasing popularity of hand held cameras and mobile phone cameras, there is a great demand for moving object tracking when the camera is moving. Unfortunately, none of the methods developed for object tracking in stationary camera conditions, are applicable for object detection or tracking when the camera is not stationary or in other words when camera is moving. This generates a great need for the development of new robust methods to track moving objects, when the camera is moving.

In this thesis work, a new method is proposed for moving object tracking in moving camera scenario. In the method proposed, first background is estimated and then background subtraction is performed. In this method, first of all geometric transformation is calculated between every two successive consecutive frames, and then we transform frame (t+1) in the plane of frame t using this geometric transformation, thus cancelling the camera motion. For determining the object motion, we have calculated motion measure (MM) and reliability measure (RM) for each pixel. This motion cue and the cue obtained from the difference of background and current frame are combined to get foreground and background mask. New background is modeled for the current frame using this foreground and background mask and then background subtraction is performed for final foreground and background labeling.

The proposed method is able to track the object present even in the first frame of the video. This method also handles fast moving background. Experimental results show that, the results are reasonably good compared to most of the other existing methods.

Table of Contents

Table of Figures:

CHAPTER 1: INTRODUCTION

In real time, the detection of moving object or target based moving object detection is very crucial research area of computer vision. Many researchers of computer vision and image processing are being attracted from decades towards object detection and tracking due to its increasing popularity. The moving object is detected using frame differencing method, background subtraction method, kernel based background subtraction, differential method, and optical-flow based methods.

1.1Overview

In most of cases, moving object may move intermittently (for example, moving vehicle at traffic light). They may not be focal point in security- surveillance camera's field of view. Sometimes unicasting background regions may exhibit cluttered [1, 2, 3] nature (e.g. spouting or flowing water, moving tree branches or leaves, flags etc.). It is a 2-class problem and because of wide range of scenarios, the some useful parameters that control adaptation rate and sensitivity can be defined by user. For the purpose of surveillance, CCTV cameras, webcam and security cameras have been installed at public places and crowded areas all over the world for security. Hence need for the development of intelligent system arises, which can be used for surveillance purpose to estimate moving objects. In moving state, the representation can be done using points, geometry, silhouette, contour, articulated shape or skeleton [4, 5]. For video analysis key steps are [5, 6, 7]:

- Moving object detection
- Frame to frame tracking of detected objects
- Object analysis

In the area of image processing in computer vision, there is currently a growing demand for the motion based object detection, which enables immersive visual surveillance and consumer electronic based security applications. Accurately extracting human regions (the foreground) from real-time or online/live video is an important feature of this work [4, 5, 7]. To strengthen the visual surveillance system for indoor outdoor applications in public and private areas, automated information detection based on motion is a demanding area of research for the purpose of security.

It is again a very challenging problem to adjust a parameter for particular problem like illumination variation, dynamic background scene and camouflage, all together present in a scene [1, 2, 8-11]. Therefore, better understanding of data and BGS method is required to achieve optimal performance.

1.2Object Detection

The Object detection in video is the most popular research area of computer vision and image processing. It involves extraction of visual information or features from sequential images or video frames with the help of digital image processing techniques. Object detection is simply the process of identifying moving object in image/video frames. Object detection handles instances of object detection of certain class in a video frame, objects may be humans, vehicles, buildings or any other objects of interest which are identifiable.

Figure 1: Object detection in video frames using bounding box

Accurately extracting human regions (the foreground) from real-time or online/live video is an important feature of this work. To strengthen the visual surveillance system for indoor outdoor applications in public and private areas, automated information detection based on motion is a demanding area of research for the purpose of security [1, 3, 4, 7]. Moving object is detected using a very popular approach called background subtraction. Various researchers has obtained foreground moving objects by modelling of background and then detect either background or foreground [3, 4, 6-8].

In computer vision, we can express background subtraction as the binary segmentation of input frame sequences. Background subtraction can also be referred as foreground detection [1, 2]. Background subtraction is useful in various vision applications for visual surveillance and security. It can be used in traffic handling and monitor, suspicious motion analysis, abandoned object identification and detection, medical applications and many other real life indoor-outdoor applications [2, 3, 4]. The advanced background subtraction approach is inherited from the traditional approach, in which the current test frame is subtracted from the background frame, which we have modelled by background modelling, and the difference frame is generated. And then the difference is compared to a threshold value to decide each pixel state as foreground pixel or as the background pixel. Practically, the achievement of the traditional approach is not good due to changes occurring in background [4, 5, 6, 7]. In literature, we have a handful of models for background subtraction, which have been developed and investigated in decades [6, 7]. But still in visual surveillance applications, this is an ever growing research area due to many challenges.

1.3Object Tracking

The moving object tracking is a vast area of research in computer vision. The rapid increase in the number of using high-powered computer systems, inexpensive video cameras, the increasing requirement for automated video surveillance and analysis has made object tracking an interesting area. In object tracking, the trajectory of one or more than one objects is followed from entry to exit point in the scene [5, 7, 9]. Object tracking is a process in which in a given video sequence, an object is followed in each frame sequentially. Generally, moving object tracking can be defined as the moving object trajectory estimation in input video frames as per the object movement in the scene. Uniform labels are given to target objects in every frame by a good tracker. Additionally, depending on the domain and application's requirement, object-centric information can also be represented by a tracker, like shape of an object or a target, area covered or moving object orientation.

In the analysis of video, tracking an object can be complex [5, 9, 10-18] because of:

- Information loss due to projection of 3D world points on a 2D image plane.
- Noise in frames due to hardware or environmental issues.
- Complex motion of object.
- Variation in the illumination of background scene.
- Very high movement of moving object.
- Non-rigid shape of moving objects.
- Complex object shapes.
- Partial occlusion of objects and full object occlusions.
- Articulated moving objects.

Figure 2: Detection and tracking of moving object [Source: J.S. Kim et. al. (2011)]

Figure 3: number plate detection, tracking and recognition in each video frame [Source: www.getti-images]

1.4Object Detection and Tracking approaches

In literature, a handful of methods have been developed for moving object detection from input video frames. Some of them are described below:

1.4.1 Background subtraction

When camera is stationary, background subtraction is a well-known technique for moving object detection. In background subtraction we obtain a mathematical model of background called as background model and this model is compared with each new input frame of the input video sequence [3-8]. In this method for moving object detection successive frames are subtracted. The main steps of background subtraction technique can be classified in three phases [12-20]:

1. Training Phase: The modelling of background reference frame has performed by trimmed mean based simple averaging method.
2. Testing Phase: In this phase, following actions have performed:
   a. Generation of suitable threshold value.

b. Difference of background reference frame and current test frame.

c. Classification of each pixel using difference frame and threshold.

3. Enhancement Phase: To improve detection quality using image processing techniques like morphological filters, boundary analysis, optimization techniques etc.

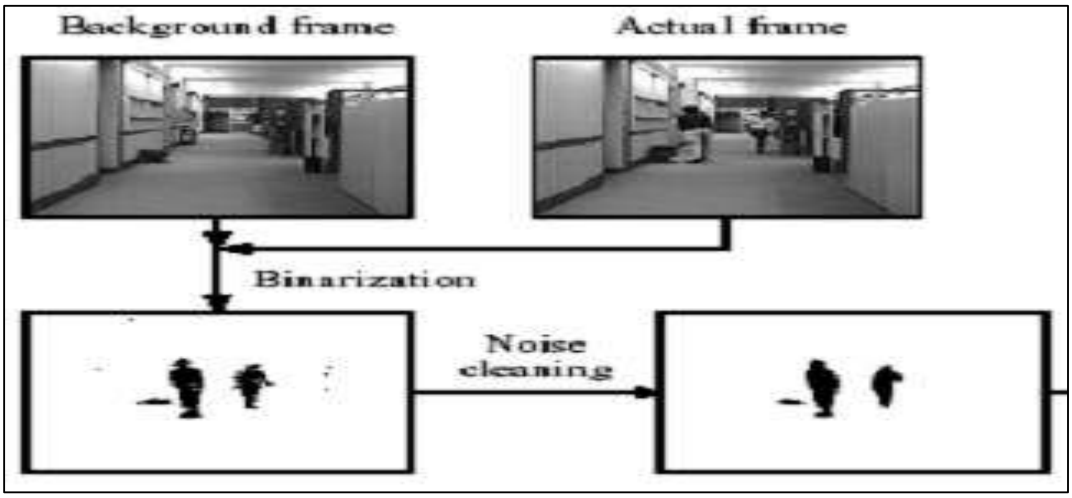


Figure 4: Object detection in video frames using background subtraction

1.4.2 Phase-Correlation method

Phase-Correlation method is a popular method, if we have to find out the relative motion between two images, in computer vision and image processing field. This method makes use of Fourier transform and inverse Fourier transform of two image frames to evaluate the correlation between them. Fast Fourier transform has made Phase-Correlation method faster than any other correlation method, and this method is useful for scaling difference determination or rotation determination between two image frames.

1.4.3 Optical flow based method

This approach is used to identify moving pixels (estimation of motion) for two consecutive frame at time t and ($t+\Delta t$). Optical flow based methods uses the brightness constancy assumption. As in case of 2D (similarly for 3D) a pixel at location (x, y) in the frame at time t with intensity $I(x, y, t)$ is moved to the new location in the frame at time (t+1) with the displacement as p in x direction, and q in y direction, then the brightness constancy assumption equation can be written as [9, 21]:

$$I(x, y, t) \approx I(x+p, y+q, t+1) \qquad \dots (1)$$

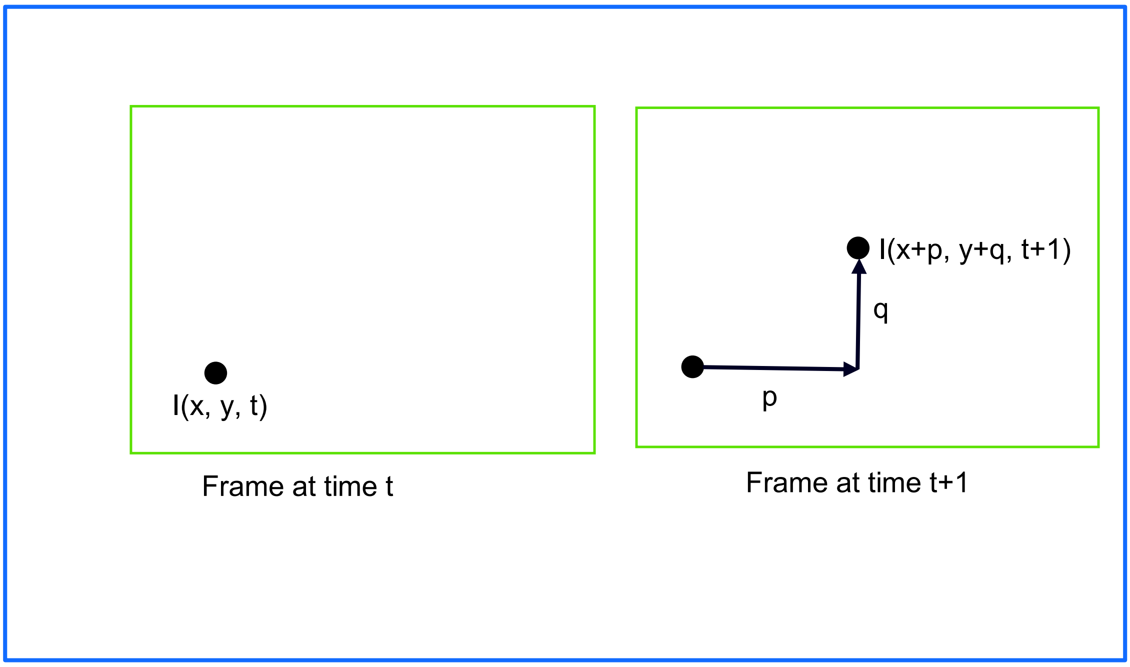


Figure 5: Brightness Constancy Assumption

The drawback of this method is the aperture problem. So the method alone is not sufficient for optical flow detection. Other equations are also required for optical flow determination, with some more additional constraints.

For the determination of optical flow between given input image frames, the differential methods are widely used and are very important methods, which have been discussed below:

- **Horn-Schunk method**

In this method, aperture problem is solved using global smoothness constraints. Flow is defined as a global energy function in this method, which is to be minimized [21]. In this method flow vectors have high density but they are unable to handle discontinuities. Noise sensitivity of this method is also high.

- **Lukas-Kanade method**

In Lukas-Kanade method, a pixel's local neighborhood has essentially constant flow. In this method optical flow equations are solved using least-square criteria for all pixels. In Lukas-Kanade method three assumptions are followed: brightness constancy assumption, spatial coherence and temporal persistence or small movements [9, 21]. If motion is large, Lukas-Kanade method does not follow the brightness constancy assumption. In case of large motion small points cannot always be found by Lukas-Kanade method, because small points can go outside of the local window and hence difficult to find out.

1.5Application Areas

In these applications, main goal is to detect the changes between consecutive frames. In background subtraction, one frame is considered as background frame or modelled background frame, also called training frame and other frame is test frame. The difference of these frames detects the changes. Using a suitable threshold value, a pixel is identified as part of foreground or background.

Figure 6: Tracking an object for security and surveillance system [Source: www.google.com]

The object detection is applicable in various kind of real life based applications [1-5, 9, 15-21] like

1. Traffic monitoring and analysis
2. Driver assistance system
3. Defence system
4. Indoor-outdoor surveillance
5. Shoplifting in retail
6. Coastal surveillances at marine
7. Air-line surveillance system
8. Border security system,
9. Police and law enforcement
10. Night vision in indoor-outdoor and traffic analysis surveillances
11. Activity analysis,
12. Abandoned object detection,
13. Logo or official sticker based vehicle detection and recognition
14. Re-identification of vehicle
15. Face detection and recognition
16. Retina detection and recognition
17. Biometric System
18. Logo detection
19. Packaging detection in packaging industry
20. Crack detection on metal sheets in manufacturing industry
21. Medical domain
22. Target based moving object detection and tracking
23. Number plate detection and tracking of moving vehicle
24. Medical diagnosis using body area network
25. Traffic monitoring and analysis using sensor network

It can be applicable in multimedia and its applications, computer vision systems, traffic analysis and monitoring, artificial intelligence and automatic computer algorithm development. In general object detection is a concept of evaluating the state of motion based pixel.

1.6Major Challenges and Issues

Apart from the above principles, in real time system there is problem of motion, illumination variation in real scene. In literature, [Toyama et.al. [1999], Theiry Bouwmans [2014], Yadav and Singh [2015] etc.] has discussed various challenges available, and these crucial challenging issues are given as-

1. Availability of noise in the image due to web cam or any transformation.
2. Camera jitter: Due to wind camera may sway back and forth that adds minor motion in sequence.
3. Camouflage: A foreground pixel's characteristic is subsumed by background pixel. This causes difficult to distinguish between background and foreground.
4. Bootstrapping: In some cases, background frames are not available, then it is very difficult to model the background.
5. Illumination variation: In many indoor-outdoor surveillance applications, illumination changes gradually or sudden, which causes various false detection present in detection process.
6. Foreground aperture: The foreground region has uniform coloured region and changes inside the foreground region caused by various illumination or motion issues, the possibility of false negative increases.
7. Motion in background: Due to dynamic nature of background like moving tree leaves, spouting or floating water or moving banner of flag in background, detects many false positive pixels.
8. Insertion of new background object: In case of new background object, there is possibility of false detection which degrades the performance of the system.
9. Sleeping foreground object: Various applications related to abandoned object detection falls in this category. In such cases a moving object becomes motionless, and then this is difficult to identify these objects as foreground.
10. Shadow detection: Detection of shadow is also an active research area. Due to various lighting issues, moving objects has its shadow that is not in the goal of object detection. It is considered as part of background. Shadow causes increase in false detection which results in degradation of robustness and performance of the system.

In literature, various datasets are available, but still there is requirement of robust method which can handle all issues. These are very critical and problematic challenging issue available in literature. There may be some more issues depending upon the problem and applications requirement in visual surveillance system. Based on above challenges and issues, proposed work can be summarized in problem statement.

1.7 Motivation

Since last some couple of years, based on available literature on background subtraction technique, now a days the background is not static, because of the motion in background, illumination change in background scene, recording is taken by handheld devices(mobile or camera) or various videos are taken from internet. Apart from dynamic background, moving camera based object detection and tracking is a new field of research. The background subtraction work on the following basic principles-

1. Camera is in static position.
2. Variation caused by illumination is constant.
3. Background scene is static.

In the background subtraction based method, the image scene can be divided into three parts-

1. Background: Part of static scene which is also visible.
2. Artefacts: Changes in the image scene due to shadow and environmental changes or light.
3. Object: Interesting portion of scene i.e. pedestrian, vehicle or any moving object.

Therefore, the development of an effective method to handle camera motion and to deal with dynamic background or illumination changes in video frames is still important. Another feature of background subtraction based background modelling methods is to develop a robust method which has to be computed in real scenario in less time and low memory requirement. Still there are following issues:

1. Lack of a common moving object detection framework.
2. Lack of significant scientific progress.

3. Absence of single realistic large scale dataset.

1.8Problem Statement

As it has been discussed above, each methodology has some advantages and also some disadvantages, hence, there is a need to pay attention towards investigation and development of novel feature extraction methodology for object detection and tracking in real time. In this work, the main motive of this research work is detection and tracking of the moving object in each video frame when camera is not stationary. That is, we will develop a method to detect and track objects which are moving in the scene; in the scenario where camera itself has some motion. In this work, we do some of the task given as:

- To develop a novel method that can extract moving object from motion based video frames in moving camera scenario.
- To identify and detect the presence of motion based object in moving video frame.
- Remove the unwanted background pixels from the detection results and hence improve the overall performance of proposed work.

1.9Expected Outcome

The proposed work outcome has to detect and track objects that are moving in scene, when the camera is also moving. It improves the shortcomings in moving object detection and tracking, which are generally present due to camera motion. Different approaches have been explored by several researchers for object detection and tracking and are being implemented. Each approach is defined for a specific problem and hence the success of a particular method is largely dependent on the problem domain. After completing all above mentioned work, the required meaningful outcomes are detected with efficient performance. The main outcomes are given as:

- To develop object detection algorithm that has fast computational speed because of frame based processing and minimal mathematical computations.
- To detect and track the moving object in moving background with dynamic scene in video.

- The proposed method has presented better performance in coloured as well as moving background with moving camera based video.

CHAPTER 2: LITERATURE SURVEY

In this chapter we discuss about moving camera Video Sequence (MCV) and background literature. The extraction of motion based foreground object from Video scenes is an important phenomenon for spatiotemporal analysis. This visual information and content representation is very useful for many real-time computer vision as well as video processing application. There are various techniques, proposed in the past based on background subtraction method for detection of moving object. In this chapter we will discuss some of the techniques in brief. Most method used for moving object detection in video frames are based on a simple idea focused on surveillance applications using stationary camera, the disparity of background model pixel and corresponding pixel belonging to current frame is mainly indicative of foreground object. The idea behind this concept is that no prior information is required to classify the state of a pixel as foreground object, their difference and suitable threshold is enough to classify the pixel state as foreground object.

In this work a brief overview of several notable research papers are considered. Detailed study of existing state of the art works and literatures are not in the scope of this thesis and therefore reader may refer to recent surveys in this area for more detailed studies [T. Boatwoman's].

Chattopadhyay and Das [1] proposed moving object segmentation from moving camera video by implementing energy minimization technique.

In this work initially,

1. An initial object cut (IOC) $c^{+}$is obtained by using a threshold then the value of saliency map can be obtined as -

   $C^{+} = \{\ C_i^{+} \mid C_i^{+} \in \{0,1\}\}$

   $C^{+} = O^{+} \cup\ B^{+}$ and $O^{+} \cap\ B^{+} = \phi$

   $O^{+} = \{o^{+} | o^{+} = 1\}$ and $\{B^{+} = C^{+} \backslash O^{+}\}$

Where O^+ is tentative foreground blob that can be obtained in frame e^+. Then applied a graph based ranking algorithm for stability.

2. In second stage, an accumulated object blob (AOB), A^+ is evaluated for t^{th} frame and can be classified as:-

A pixel a_i^+ of an AOB having coordinate (x_i, y_i) is labeled as -

$a_i^+ = f_{majority}[e^{(t+k)}(x_i, y_i)$, k = -floor $[\frac{w}{2}]$, …., floor $[\frac{w}{2}]$

Where w<<n and w is size of sliding window.

$f_{majority}$= {1 if majority of set of pixels is 1

0 if majority of set of pixels is 0

3. In the third stage, this work proposed a trimap generation model that evaluates the strongness between foreground and background pixels.

4. In this stage, Chattopadhyay et. al. proposed a graph cut based optimization technique for energy minimization.

This model reduces the error by incorporating past and future records for processing as well as updating the model. This work tested on various challenging video (MCV) scenes but sometime it was not able to detect whole object.

A research work proposed by [2] developed a combined approach of temporal and spatio-temporal histogram based background subtraction. This work also experimented over various problematic MCV shots. This work improves the detection quality by overcoming the errors of considered state of the art models. The abrupt illumination variation is handled by median filter. This model adopts a histogram based representation of pixel intensity distribution.

S.W. Kim et. al. [3] developed a background model for moving camera video frames that is suitable in various different conditions like illumination variation, fast movement of camera. This method also handles issues caused by Zooming of video sequences. This work applied a 2D DOG filter for sharpening of video frames that strengthens the edges of object. The general equation of 1 D DOG can be expressed as-

$$f(x, \mu, \sigma_1, \sigma_2) = \frac{1}{\sqrt{2\Pi}}[\frac{1}{\sigma_1}\exp(\frac{-(x-\mu)^2}{2{\sigma_2}^2}) - \frac{1}{\sigma_1}\exp(\frac{-(x-\mu)^2}{2{\sigma_2}^2})] \quad \dots (1)$$

In this work false alarms are eliminated by considering neighborhood pixels in the decision process. For each pixel X_e of current frame, the candidate anchor point X_b is evaluated from background model and computed as -

$$X_b = \text{H}.X_e \qquad \ldots (2)$$

Where H is a transformation matrix.

In this work, Kim et. al. [3] developed a model for labeling of each pixel X_e as foreground and background and is classified by comparing X_e to the neighbourhood pixel of $X_b$from N (X_b)as -

$$D_{X_e,X_b} = \frac{1}{\sigma^2(X_i)}(I(X_e) - \mu(X_i))^2 \qquad \ldots (3)$$

$$cap\ X_b = arg\ min_{X_i} D_{X_c,X_i}, \text{ where } X_i \in N(X_b) \qquad \ldots (4)$$

and

$$\text{L}(X_e) = \{ \text{ BG if } D_{X_c,cap\ X_b} < T_c; \quad \text{FG otherwise} \qquad \ldots (5)$$

Where, σ^2 is SD and μ is mean of BGM and I (X_e) is intensity of current frame. CapX_b is a point in the BGM that correspond toX_c. The threashold value T_c is used to decide the labels of each pixel.

Vaikole and Sawarkar [4] proposed semi-automated method for object detection and segmentation that can be used in many computer vision applications. This work resolves the issues generated during moving camera and uncovered background. In case of moving camera, a motion estimation based method and in case of uncovered background a simple registration technique has been used. In this work temporary poses has been handled by integrating the region based segmentation.

The proposed research work of Kim and Kweon [5] has shown an object detection and tracking model for visual surveillance under motion based camera. In this work, object is detected by applying Kanade-Lukas-Tomas [KLT] features tracker on two consecutive video frames and an online-boosting tracker is used to track those objects in sequential frames. This work detects and track moving object without background modeling.

The key features of this work can also be expressed as:-

1. Homography – transformation based motion detection.
2. Online Boosting Tracker.
3. Integration of detector and tracker.

In this research work, A. Ghosh et. al. [6] proposed an object or foreground detection method from video sequences captured by camera in moving scenario. This work proposed a combined approach of edge incorporated MRF (Markov Random Field) with local histogram based matching. The Fuzzy edge incorporated MRF is mainly used for spatial segmentation. Parameters in the MRF model have been estimated through EM algorithm. Here, A χ^2 test is applied for local histogram based matching for moving object detection in complex scene.

In [7], Jin et. al. developed a multi-layer Homography based algorithm for modeling of background where video is captured from free-moving camera.

In the modeling process, the probability of each pixel to the part of background is evaluated as-

$$\mathrm{P}(I^{(t)}(x^{(t)})) = \sum_{i=1}^{K} w_i^{(t)} \, (X_{l_h}^{(t)} . \eta(I^{(t)}(x^{(t)})), \mu^{(t)}(x_{l_h}^{(t)}), \epsilon^{(t)}(x_{l_h}^{(t)})) \qquad \ldots (1)$$

Where k is no. of Gaussian and the Gaussian density function is described as:

$$\eta(I, \mu, \Sigma) \; = \; \frac{1}{(2\Pi)^{\frac{h}{2}} |\Sigma|^{\frac{1}{2}}} exp(\frac{-1}{2}(I - \mu)^T \Sigma (I - \mu) \qquad (2)$$

The likelihood of each pixel is defined as:-

$$l(l(x) \mid I(x)\,) \; = \; P(I(x)\,)\; ; \; l(x) \; = \; 0$$
$$1 - P(I(x)\,)\; ; \; l(x) \; = \; 1 \qquad (3)$$

Where w is weight parameter and $l(x)$ is label and label $l(x)$=0 denote the background and $l(x)$=1 denote the foreground.

In this chapter, we have explored some other state-of-the-art methods and are discussed as:

Irani et al. (1994) [8] proposed a method to detect and track the transparent and occluding moving objects. They proposed to evaluate the pixel state based on motion measures and reliability measures, based on the score they categorized a pixel between moving and stationary. Sharpness of the tracked object is being maintained by temporal integration while comparison of input frame with internal image representation allows the motion analysis algorithm to continue tracking of same object in further frames and thus improves the segmentation. Waving tree leaves and variation in intensity problem has not been handled in this method, so this may produce noise. The method is able to detect shadow also but not able to eliminate it.

Stauffer and Grimson [9] proposed a GMM based adaptive probabilistic for background model for moving object detection using background subtraction where every pixel was modeled with mixture of Gaussians. Main motive of this work is to classify a pixel as background or foreground. In this model the probability of observing a pixel $f_t(x, y)$ is

$$P(f_t) = \sum_{k=1}^{K} w_{k,t} * \eta(f_t, \mu_{k.t}, \varepsilon_{k,t})$$

$$\eta(f_t, \mu_{k,t}, \varepsilon_{k,t}) = \frac{\exp(-\frac{1}{2} * (f_t - \mu_{k.t})^T \varepsilon_{k,t} (f_t - \mu_{k.t}))}{\pi |f_t| |\mu_{k.t}|^{1/2}}$$

Where,$w_{k,t}$: weight of Kth Gaussian component.

$\mu_{k.t}$: mean of Kth Gaussian and

$\mu_{k.t} = \sigma_{k,t}{}^2 * I$, where σ^2 is variance and $\varepsilon_{k,t}$ is co-variance matrix.

Parameter of Gaussian distribution are updated for each pixel, if a match is found for pixel $f_t(x, y)$ within 2.5 times of standard deviation of kth Gaussian distribution then

$$w_{k,t} = (1 - \alpha) * w_{k,t-1} + \alpha$$

$$\mu_{k.t} = (1 - \rho) * \mu_{k.t-1} + \rho * f_t$$

$$\sigma_{k,t}{}^2 = (1 - \rho) * \sigma_{k,t-1}{}^2 + \rho * (f_t - \mu_{k.t})^T * (f_t - \mu_{k.t})$$

ρ is the Gaussian pdf and scaled by parameter α

$$\rho = \alpha * \eta(f_t, \mu_{k,t}, \varepsilon_{k,t})$$

If a pixel is non-moving and part of background, then do not go far from $|f_t - \mu| < 2.5$ i.e. for unmatched Gaussian distribution, the mean and variance do not change. In this case, weight is updated. Finally, a pixel is classified as background if $arg(min(\sum_b^K w_b\ ))$ where T is threshold used to decide what portion of data will be part of background.

M. Haque et. al.[10] proposed a GMM based background model using BBS. This model removes the dependency on rate of learning (α) and threshold T used by [S&G] because [S&G] is very sensitive to the noise with varying learning rate. Therefore, this model [M.Haque] eliminates T and replaces T with another parameter that improves the robustness of this technique. In this model K-Gaussian components are used with meanμ_k, variance $\sigma_k{}^2$ and weight w_k such as $\sum w_k$ = 1 for each iteration Gaussians are ordered by $\frac{w}{\sigma}$ in descending order. The variance of moving foreground pixel is more as compare to the static background pixel. For each pixel f_t at time t, first match the scatterness of model for new pixel f_t . If $|f_t - \mu_k| < 2.5\ \sigma_k$ then

$$\mu_k = (1-\alpha)\mu_k + \alpha f_t$$
$$\sigma_k{}^2 = (1-\alpha) * \sigma_k{}^2 + \alpha * (f_t - \mu_k)^2$$
$$w_k = (1-\alpha)w_k + \alpha$$

And the weights of other components are updated as

$$w_k = (1-\alpha)w_k$$

Otherwise, introduce a new Gaussian with initial values μ $= f_t;\ \sigma_{init} = 30;\ w_k = w_{init} = 0.001$.

In [9] model Gaussian components are used to model background but here in [6] considered only most dominating Gaussian component to model the background. If I_t is the pixel value of most dominant Gaussian component and f_t is the current pixel value then the pixel is considered as background if $|f_t - I_t|$ < S otherwise pixel is considered as foreground. Where S is low and low value of S guaranties better quality of moving object detection. This model handles shadow and reflection of light very well.

The main limitation of above mentioned GMM based background subtraction technique is that it does not work well in the following cases:

1. Changing surface of water (with sunlight and shadow) or fast motion in background.

2. Not suitable for sudden illumination changes in background. In this case foreground pixels merged into background and remain as part of background for long period of time.

C R Jung [11] proposed to model the background using statistical descriptors and extract foreground pixels. And then geometrical constraints are used to remove shadows. Outliers in the generated distribution by moving object in the scene is handled using robust mean estimator. α-metrically trimmed mean (λ_{α}) is used to develop BG model.

$$\lambda_{\alpha}(x) = \frac{1}{T - [\alpha T]} \sum_{t \in S_{\alpha}(x)} f_t(x)$$

Where $S_{\alpha}(x) = \{t : f(x,t) \leq T - [\alpha T]\}$

With the help of scale parameter σ, the noise spread is estimated around the actual background.

$$\sigma(x) = 1.4826 * \mathrm{MAD}(x)$$

Where, $MAD(x) = median_{t \in \{1,2,\ldots T\}}\{|f_t(x) - M(x)|\}$

The value foreground pixel is for from the mean value in distribution. A pixel is foreground pixel if

$$\textstyle\sum_{u \in \Lambda(x)} w(u)\,|f_t(u) - \lambda(u)| > K \sum_{u \in \Lambda(x)} w(u).\sigma(u)$$

Where, $\Lambda(x)$: 3x3 pixel neighborhood centered at x.

w (u): weighted mask at each pixel.

K: parameter to handle deviation from mean with respect to σ.

Holes which are present in interior part of object can be filled by morphological operations (such as closing and opening) and isolated pixels are also removed from the foreground. A pixel ratio based shadow was developed to handle shadow.

- One of the limitations of this model is that number of false positives (where shadow pixels are considered as object pixels) have generated.
- If the background is homogeneous and the object present in front of this homogeneous background is also homogeneous then, the false detection increases in this method.
- This approach fails if luminance component of foreground pixels are similar or exactly same although foreground pixels have different colors. That is, this methodology does not rely on luminance.

Akula et. al.[12] proposed a method for object detection in infrared frame sequences using background subtraction technique. This method is based on adaptive contour. They have classified their work in three stages:

a. Foreground object detection using statistical background subtraction

$$F(x,y) = \begin{cases} 1, \dfrac{|I(x,y) - \mu(x,y)|}{\sigma^2(x,y)} > T^2 \\ 0, otherwise \end{cases}$$

Where I(x, y), μ(x, y) and σ^2(x, y) are pixel value of current test frame, mean value and variance of BG model frame respectively.

b. Contour detection using contour saliency map (CSM)

The CSM based contour model is developed using –

$$CSM = \min(\frac{\|(I_x, I_y)\|}{Max}, \frac{\|(I_x - B_x), (I_y - B_y)\|}{Max})$$

Where I_x, I_y are gradients of test frame in x and y direction respectively.

B_x and B_y are gradients of BG modelled frame in x and y direction respectively and Max is normalization factor. The value of CSM lies in range of [0, 1] and larger value represent stronger confidence and validates that the particular pixel belongs to the boundary of moving objects. The BG model is updated adaptively with each test frame and learning parameter α.

Tiwary et. al. [13] proposed a background subtraction based hybrid approach for moving object detection in multiresolution based framework for thermal based infrared video sequence. This work developed a fast ICA based background subtraction method, BG and optical flow [proposed by Horn and Schunk, 1981] based hybrid approach for object detection and updates BG adaptively. The main focus of this work is to handle camouflage conditions. This work will perform under high ambient temperature in summer and cold ambiences in winter days and hence suitable to cover wide range of ambience of temperature variation.

Zhou et. al. developed a colour invariant based BGS based methods for real time surveillance video sequences of indoor environment where FG moving object and still BG both have similar color.

A difference of GMM (DoG) based method proposed by X. Dong et. al. [14] has introduced a novel moving target detection method in thermal video frames under complicated BG. This work extracts interest point with DoG filters then tracks these interest points for each frame and finally R-means clustering techniques classify these interest points into FG or BG. The 2D Gaussian function is expressed as

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{\left(\frac{-(x^2+y^2)}{2\sigma^2}\right)}$$

Then the difference of two Gaussians –

$$DOG(x, y, \sigma_1, \sigma_2) = G(x, y, \sigma_1) - G(x, y, \sigma_2)$$

$$= \frac{1}{2\pi} [\frac{1}{\sigma_1{}^2} e^{\left(\frac{-(x^2+y^2)}{2\sigma_1{}^2}\right)} - \frac{1}{\sigma_2{}^2} e^{\left(\frac{-(x^2+y^2)}{2\sigma_2{}^2}\right)}]$$

After applying difference of Gaussian filter, a feature map group is generated with distinct scales, then by combining these features maps, a saliency map is generated as given below:-

$$Sali(x, y) = \max_{\sigma_1, \sigma_2} \{I(x, y) * DOG(x, y, \sigma_1, \sigma_2)\}, \qquad \sigma_1 < \sigma_2$$

Where I is input frame and '*' represents convolution operation. The interest point (x, y) is identified as:-

$If\ Sali(x_1^n, y_1^n) > th, \qquad n = 1,2, \ldots \ldots N\ N\ is\ total\ number\ of\ interest\ point (IP)$

Then $IP_1 = \{x_1^n, y_1^n | Sali(x_1^n, y_1^n)\ is\ local\ maximum\}$

$$IP_1^n => represents\ position\ of\ nth\ interest\ point\ in\ first\ frame$$

Now track these interest points and R means algorithm is used to cluster the IP.

B. Bondzulic proposed a multi sensor BGS based hybrid approach for visual surveillance system. The multi sensor based systems aimed to captures several images with different modality of same scene. By combining the information of each source, a robust multi sensor based system was developed for visual surveillance system.

M. Teutch et. al. developed a person detection method with moving thermal camera in low resolution i.e. thermal long wave infrared camera (LWIR). This approach detects HOT

SPOT instead of BGS and a DCT based descriptor verified hot spots and a modified random naive Bayes (RNB) classifier is used for classification.

V. Reddy et. al. [15] has given a block based method for foreground detection using probabilistic decision method that is capable of dealing with problematic issues like dynamic background, noise, illumination variations.

This work can be briefly summarised in four stages-

1. Overlapping blocks are generated by dividing frame then a low dimension descriptor was generated for each block.

2. The classification of each pixel of every block into background or foreground using GMM and cosine distance transform.

- A multivariate GMM is used to identify the state of a pixel in a block as background and able to handle dynamic background like waving tree, fountain, spouting water etc.

$$p\big(d_{(i,j)}\big) = \frac{exp^{\{-\frac{1}{2}[d_{(i,j)}-\mu_{(i,j)}]^T \Sigma_{(i,j)}^{-1}[d_{(i,j)}-\mu_{(i,j)}]\}}}{(2\pi)^{D/2}\left|\Sigma_{(i,j)}\right|^{1/2}}$$

$$and\ if\ p\big(d_{(i,j)}\big) \geq T_{(i,j)}$$

  Then all pixels of a block = Background

  Where $T_{(i,j)} = p\big(t_{(i,j)}\big)$ and $t_{(i,j)} = \mu_{(i,j)} + 2diag(\Sigma_{(i,j)})^{1/2}$

  And $d_{(i,j)}$is 12 dimensional descriptor, using diagonal covariance matrix, thethreshold value covers approx. 95% of the distribution.

- The cosine distance transform is used to resolve the issue of illumination variations like sudden/gradual light change etc. By applying the cosine distance based classifier, block (i, j) is classified as background part.

$$cosdist\big(d_{(i,j)}, \mu_{(i,j)}\big) = 1 - \frac{d_{(i,j)}^T \mu_{(i,j)}}{\left\|d_{(i,j)}\right\| . \left\|\mu_{(i,j)}\right\|}$$

The next classifier based on temporal correlation (mostly exist in frame sequences) used to minimize false positive detections caused by various changes in environment. In this step, a block (i, j) is classified as part of background if these conditions are met:

(1) $d_{(i,j)}^{[prov]}$Was classified background previously.

(2) $cosdist\left(d_{(i,j)}{}^{[prov]}, \mu_{(i,j)}\right) \leq C_2$

Where $C_2 = 0.5 * C_1$

In this work, each of the three classifiers faces a distinct issue or situation such as varying illumination or dynamic nature of background scene. The third classifier aimed to minimize false positive detection by analysing temporal correlations between consecutive frames. In this work the state of the art models classify pixels into background or foreground whereas the proposed work makes decisions on blocks that is group of pixels.

Zhou et. al. [16] proposed a framework for "DEtecting Contiguous Outliers in the LOw rank Representation (DECOLOR)" in video sequence. The main contribution of this model is to detect moving object and BG simultaneously without training sequence and can be discussed as:-

In this model, a matrix composed of vectorised frame sequence is approximated by a (RPCA based) matrix having low rank property and the outliers in this low rank matrix can be represented as moving object.

In video sequences, the moving objects are usually shown in small clusters. So the outliers support is explicitly modelled for contiguous regions. This can be done using Markov Random Fields (MRFs) which is proposed by [16]. The motion of BG seen due to camera motion, a parametric motion model is used to resolve this issue. This model has some merits and demerits. As it avoids the complicated motion by proposing outlier detection and make low rank modelling suitable for dealing with complex BG. The DECOLOR work in batch mode, therefore, not suitable for object detection in real time system.

Haines et. al. [17] proposed a BGS method using Dirichlet process Gaussian mixture model for estimating per pixel BG distribution for both, colour and thermal video sequences. This method was unable to handle sudden complex light change in indoor video surveillance. Because of GMM this model is able to handle dynamic BG and detect moving object, having dynamic BG very well.

ViBe model [18] introduced the concept of pixel based BG subtraction complexity factor in order to estimate the speed of Visual Background Extractor algorithm proposed by Droogenbroick et al in [ViBe]. The main steps of ViBe model involved as distance measure between BG model frame and current input frame, distance based threshold value for classification of pixel. The measure contributions are counter checks, memory substitution in memory less updating policy and spatial diffusion technique for object detection in video frame sequences.

Charles et. al. [19] presented a universal method for video surveillance using spatio-temporal binary similarity descriptors and mainly detect suitable local changes due to camouflaged FG objects. The sensitivity and adaptation speed is controlled dynamically using feedback scheme. The proposed pixel label modelling {with LBSP (Local Binary Similarity Pattern) features} allows high speed implementation in parallel. The colour-LBSP based BG model B has been constructed as B(x) = {B1(x), B2(x) … Bn(x)}. The pixel classification has been presented as

$$\mathrm{S}_t(\mathrm{x}) = \begin{cases} 1, & if \quad \in \{dist(\mathrm{I}_t(\mathrm{x}), \mathrm{B}_n(\mathrm{x})) < R, \quad \forall n\} < \in_{min} \\ 0, & otherwise \end{cases}$$

Where S_t is resulting segmentation map for output.

$dist(\mathrm{I}_t(\mathrm{x}), \mathrm{B}_n(\mathrm{x}))$: Distance between current observation and background.

R: Max. distance threshold

$\in_{min}$: minimum number of matches for classification of background.

$\in_{min}$ = 2 is used, N = 20 is used.

The color LBSP distance $D_{min}(x)$ has been proposed as:

$D_{min}(x) = D_{min}(x)(1-\alpha)+D_t(x).\alpha$

Where α : learning rate,

$D_t(x)$: minimal normalised color –LBSP distance between B(x) and $I_t(x)$

$$\mathrm{D}_{min}(\mathrm{x}) \approx [0,1] \text{and } \mathrm{D}_{min}(\mathrm{x}) = \begin{cases} 1, & foreground \\ 0, & background \\ & else \end{cases}$$

and high value of $D_{min}(x)$ represents foreground objects.

This model also handle blinking pixel, which new in the literature. The blinking pixels handled by using XOR operation

$$V(x) = \begin{cases} V(x) + V_{incr};\ X_t(x) = 1 \\ V(x) - V_{decr};\ otherwise \end{cases}$$

Where $V_{incr} = 1$; $V_{decr} = 0.1$; and V(x) always ≥ 0.

The proposed algorithm has been commercialized successfully and systematically represented as a combination of multiple techniques as proposed results provides successful result in actual application.

Zamalieva et al. [33] proposed motion based estimation to model camera motion. They have combined motion, appearance, temporal and spatial cues in MAP-MRF optimization framework. Both the homography transform and fundamental matrix are used for background motion estimation. First of all for each new frame feature point is calculated using optical flow. Then matching features are used to calculate homography transform and fundamental matrix between frames using RANSAC. Geometric Robust Information Criterion (GRIC) is used for Bayesian selection of the two transformations. The geometric model with lowest GRIC score is used for appearance model propagation between consecutive frames. If Homography transform is chosen the mapping is 1-1. If fundamental matrix is chosen based on lowest GRIC score then the mapping becomes one-to-many. And a point x in frame t, maps to an epi-polar line in background. When fundamental matrix is used, the background contains more than one plane. Each of which is transformed using different homography transform, that are calculated using cascade of RANSAC steps. Homography transform and fundamental matrix together takes into account all possible frame to frame transformation. But the motion information alone isn't sufficient for accurate result. So appearance modelling, spatial and temporal cues are also used to give accurate result. Appearance modelling is done using GMM. For temporally and spatially consistent result, temporal and spatial smoothness are also considered as cost function. When motion, appearance, temporal and spatial smoothness terms are known final background and foreground labelling is done by minimizing the cost function which is energy minimization and achieved using graph-cut algorithm.

Ali Elqursh et al. [34] proposed a method in online framework for background subtraction by a freely moving camera using long term pixel trajectories. This method can handle arbitrarily long sequences. They have used Bayesian filtering framework along with long term pixel trajectory to estimate appearance and motion models. The low dimensional representation of the long term trajectories is updated automatically. By using Bayesian filtering and motion and appearance cues, different labelling belief is maintained rather than single output of segmentation. Their method has several advantages compared to state of the art methods. It handles arbitrarily long sequences and hence have long term motion information. Hence it can easily discriminate different objects with different motion. It can be used for tracking multiple moving objects and also have their appearance and motion information. It can handle partially occluded objects by appearance information of foreground and background.

CHAPTER 3: PROPOSED METHOD

3.1 Geometric Transformation

3D world points captured from camera from different positions and angles, in 2D image plane are related by a geometric transformation called as homography. Using geometric transformation an image in a plane can be defined several times in any other plane with a global co-ordinate system. A point is represented by a vector in space, in many linear algebra applications. A set of vectors defines a geometric object. Geometric transformations are nothing but algebraic operations that can translate, rotate or deform an object. Each geometric transformation inhabits some invariance properties. These properties are useful characteristics of the transformations that do not change. Linear transformations preserve straight lines.

Translation 2D:

$$X = x + \delta x,\ Y = y + \delta y$$

$$\text{In matrix form,} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} \delta x \\ \delta y \end{bmatrix}$$

Here, X and Y are not linear combinations of x and y.

Rotation 2D:

$$X = x\cos\theta - y\sin\theta,\ Y = x\sin\theta + y\cos\theta$$

$$\text{In matrix form,} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$x = r\cos\varphi,\ y = r\sin\varphi,\ X = r\cos(\theta+\varphi),\ Y = r\sin(\theta+\varphi)$$

Therefore, X and Y are linear combinations of x and y.

But in translation, X and Y are not linear combinations of x and y. Hence, it is not possible to replace translation and rotation transformation sequence by product of 2×2 matrices. But to produce K rotations, we can take product of k rotation matrices. The solution of this problem is homogeneous coordinates. In homogeneous coordinate translation and rotation are defined as:

Translation 2D:

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \delta x \\ 0 & 1 & \delta y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Rotation 2D:

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} cos\theta & -sin\theta & 0 \\ sin\theta & cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

3.1.1 Hierarchy of geometric transformations:

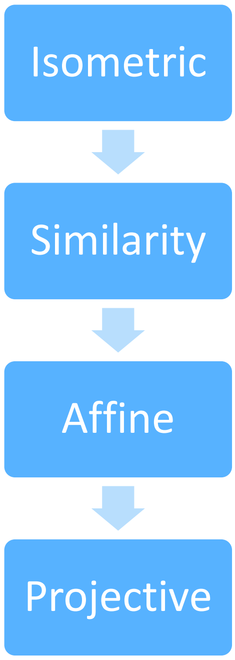


Figure 7: Hierarchy of geometric transformation

1. **Isometry Transform:**

   It is a rigid transformation. It preserves Euclidean distance. It includes translation, rotation and reflection. In Isometric transformation X = Hx is defined as:

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} cos\theta & -sin\theta & t_x \\ sin\theta & cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} => $$
$$X = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} x$$

R is an orthogonal matrix. It has 3 degrees of freedom (2 translation, 1 rotation).

2. **Similarity Transform:**

It is a set of transformations which changes the object size. It includes dilation (shrinking and enlargement) along with translation, rotation and reflection. It preserves lines, planes, parallelism, and angle between two points. It preserves the shape of the object but not necessarily the orientation. In Similarity transformation X = Hx is defined as:

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} scos\theta & -ssin\theta & t_x \\ ssin\theta & scos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathrm{x} \\ y \\ 1 \end{bmatrix} => $$

$$X = \begin{bmatrix} sR & T \\ 0 & 1 \end{bmatrix} x$$

R is an orthogonal matrix. It has 4 degrees of freedom (2 translation, 1 rotation, 1 scaling).

3. **Affine Transform:**

   It is a combination of translation, rotation, magnification and shear. It preserves parallelism also known as collinearity and ratio of distances. In Affine transformation X = Hx is defined as:

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & t_x \\ a_3 & a_4 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} => $$

$$X = \begin{bmatrix} A & T \\ 0 & 1 \end{bmatrix} x$$

A is non-singular. It has 6 degrees of freedom (2 translation, 1 rotation, 1 scaling, 1 scaling direction, 1 scaling ratio).

4. **Projective Transform:**

   Projective transformation is also known as homography. In this transform cross ratio is preserved. In Projective transformation X = Hx is defined as:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} \mathrm{x} \\ y \\ \mathrm{z} \end{bmatrix}$$

It has 8 degrees of freedom.

3.1.2 Estimation of geometric transformation (homography) between two image pairs:

The purpose to find out geometric transformation between two image pairs is image registration. If two images taken from camera are from different positions and angles, then the two images are linked by a homography. With the help of homography the two images can be transformed in the same plane and the effect of camera motion may be cancelled. Image registration has applications in the field of medicine, entertainment, remote sensing etc. To do image registration basic steps are;

- Feature detection and interest point extraction
- Feature matching
- Calculate homography matrix H based on matched feature points using MSAC (M-estimator SAmple and Consensus) algorithm.
- Use the homography matrix to align both images using

$$X = Hx \text{ or } \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

3.2 Background Subtraction

In real time, the detection of moving object or target based moving object detection is very crucial area of research in image processing. It has various applications like indoor outdoor visual security surveillance, target detection, traffic monitoring, traffic analysis, activity analysis, abandoned object detection, logo or packaging detection in packaging industry etc. Moving object is detected using a very popular approach called background subtraction. Various researchers has obtained foreground moving objects by modelling of background and then detect either background or foreground. In these applications, main goal is to detect the changes between consecutive frames. In background subtraction, one frame is considered as background frame or modelled background frame, also called training frame and other frame is test frame. The difference of these frames detects the changes. Using a suitable threshold value, a pixel is identified as part of foreground or background.

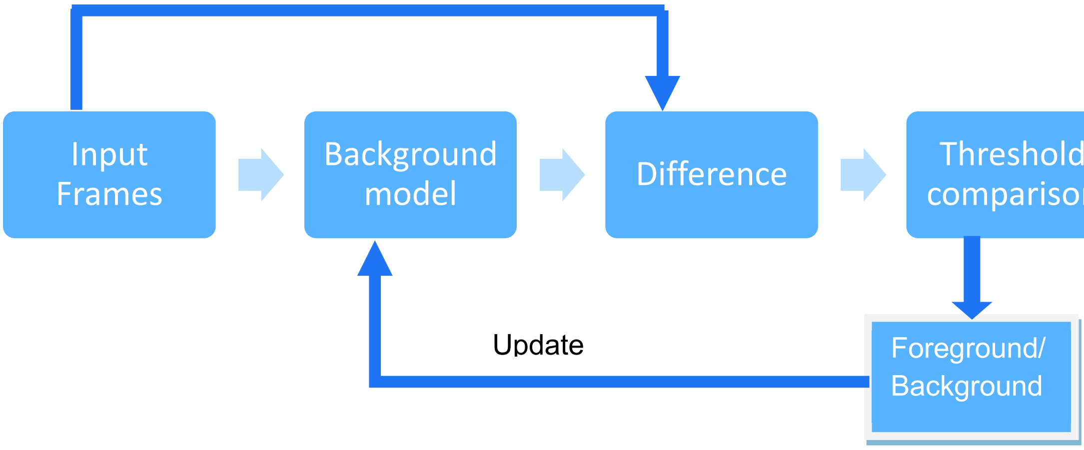


Figure 8: Block diagram of background subtraction

3.3 Methodology

We will discuss the proposed methodology in this section for moving object (Foreground) detection in moving camera scenario. Here we can notice that there are two motions in this problem:

i. Object motion
ii. And Background motion

The problem is to detect object motion not the background motion. To start with here we assume that in given scenario object motion will be much greater than the camera motion i.e.

Object motion >> Camera motion.

Hence object motion is the dominant motion. For moving object detection and tracking we have to find out this dominant object motion. To calculate object motion, we first have to cancel the camera motion.

We may assume that due to camera motion the background pixels between two consecutive frames may translate, rotate, or project on some other plane. In other words background pixels between two consecutive frames may be related with some homography matrix (H), as we know that 3D world points taken from camera from different positions and angles in 2D image planes are linked by a homography H.

$$P'_{3\times1} = H_{3\times3}P_{3\times1}$$

$$\begin{bmatrix} X' \\ Y' \\ 1 \end{bmatrix} = [H] \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

We first start with the estimation of homography matrix H, using optical flow based feature extraction and matching and using MSAC algorithm. Then we transfer the $frame_{(t+1)}$ in the plane of $frame_{(t)}$ using this homography, thus cancelling the camera motion.

After cancellation of the camera motion our next challenge is to detect the moving object or foreground object in the frame. This has been achieved in two steps:

i. Background estimation for the current frame
ii. Apply background subtraction to get the foreground

3.3.1 Background Estimation

For background estimation for the current frame, we use following steps:

i. Estimate the Motion Measure (MM) and Reliability Measure (RM) between two consecutive frames.
ii. Categorize the pixels as moving, stationary, or unknown using MM and RM.
iii. Apply Background Subtraction between previously estimated Background and Warped current frame.
iv. Combine the cues of Motion based estimation and BGS based estimation to get the Foreground mask and background mask for the current frame.
v. Estimate the Background for the current frame based on Foreground and Background Mask.

We have used optical flow brightness constancy assumption for motion estimation. Let a pixel at location (x, y) in the frame at time t with intensity I(x, y, t) is moved to the new location in the frame at time (t+1) with the displacement as p in x direction, and q in y direction, then the brightness constancy assumption equation can be written as [9,21]:

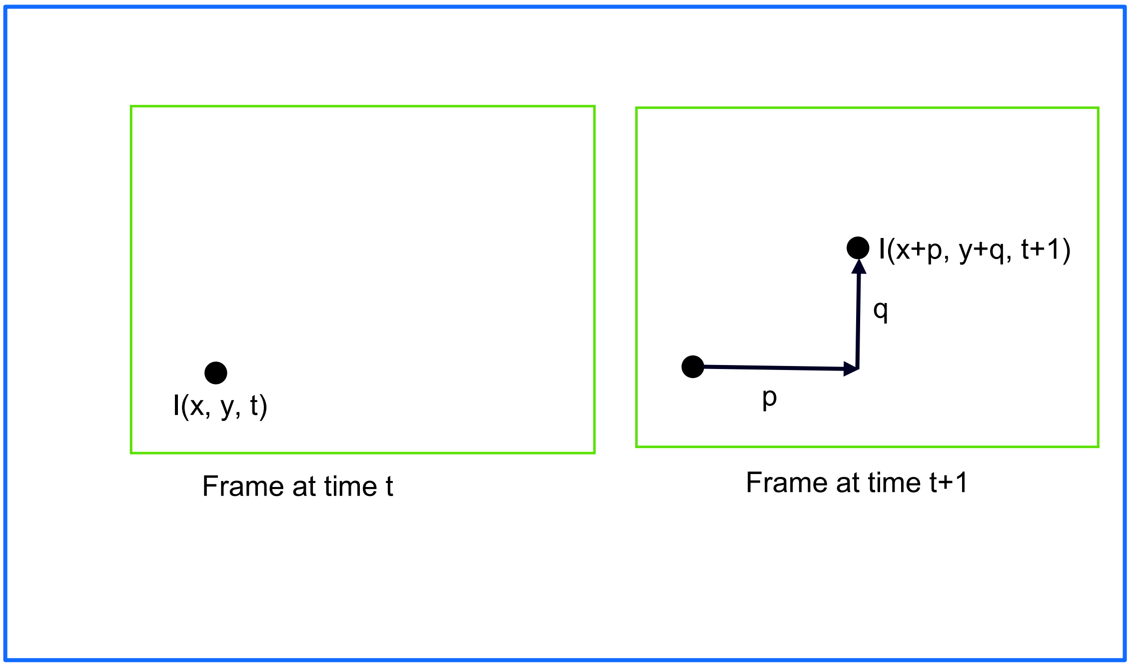


Figure 9: Brightness constancy between frame t and t+1

$$I(x+p, y+q, t+1) \approx I(x, y, t) \qquad \text{……….(1)}$$

I(x+p, y+q, t+1) is expanded using first order Taylor series expansion with respect to (x, y, t).

$$I(x+p, y+q, t+1) = I(x, y, t) + PI_x + qI_y + I_t + \text{higher order terms} \qquad \text{…….. (2)}$$

Where

$$I_x = \frac{\partial I(x,y,t)}{\partial x}$$

$$I_y = \frac{\partial I(x,y,t)}{\partial y}$$

$$I_t = \frac{\partial I(x,y,t)}{\partial t}$$

From equation (1) and (2)

$$\mathrm{P}I_x + \mathrm{q}I_y + I_t = 0 \qquad \text{…………… (3)}$$

Motion (p, q) should be such that error function is minimized at frame t. We define error function as

$$f(t)_{(p,q)} = \sum_{(x,y)\in N}(\mathrm{P}I_x + \mathrm{q}I_y + I_t)^2 \qquad \text{………(4)}$$

Motion (p, q) should be such that error function is minimized at frame t. Hence

$$\frac{\partial E}{\partial p} = 0 \; and \; \frac{\partial E}{\partial q} = 0 \qquad \text{….......(5)}$$

Let P = t_x and q = t_y

Therefore, $f(t)_{(p,q)} = \sum_{(x,y)\in N}(t_x I_x + t_y I_y + I_t)^2$

Differentiating error function f with respect to t_x and using equation (5):

$$\frac{\partial f}{\partial t_x} = 2 \sum_{(x,y)\in N} (t_x I_x + t_y I_y + I_t) I_x = 0$$

$$(\textstyle\sum_{(x,y)\in N} I_x)^2 t_x + (\textstyle\sum_{(x,y)\in N} I_x I_y)\, t_y = -(\textstyle\sum_{(x,y)\in N} I_t I_x) \qquad \text{……. (6)}$$

Differentiating error function f with respect to t_y and using equation (5):

$$\frac{\partial f}{\partial t_y} = 2 \sum_{(x,y)\in N} (t_x I_x + t_y I_y + I_t) I_y = 0$$

$$(\textstyle\sum_{(x,y)\in N} I_x I_y)\, t_x + (\textstyle\sum_{(x,y)\in N} I_y)^2 t_y = -(\textstyle\sum_{(x,y)\in N} I_t I_y) \qquad \text{……. (7)}$$

Writing equation (6) and (7) in matrix form representation

$$\begin{bmatrix} (\sum_{(x,y)\in N} I_x)^2 & (\sum_{(x,y)\in N} I_x I_y) \\ (\sum_{(x,y)\in N} I_x I_y) & (\sum_{(x,y)\in N} I_y)^2 \end{bmatrix} \begin{bmatrix} t_x \\ t_y \end{bmatrix} = \begin{bmatrix} -(\sum_{(x,y)\in N} I_t I_x) \\ -(\sum_{(x,y)\in N} I_t I_y) \end{bmatrix} \qquad \text{…... (8)}$$

Moving pixels and stationary pixels in an image pair are identified using local normalized differences. We have used local normalized differences for classification of pixels as moving and stationary, and not simply grey level differences between image pairs. Grey level differences between image pairs are not enough for classification of pixels as moving and stationary because of two reasons:

1. Uniform intensity regions could be identified as moving or as stationary.

2. Gradient magnitude in the direction of movement is also a factor affecting change in intensity due to motion.

Therefore, instead of using simple grey level differences for motion classification we have used local normalized differences between image pair to calculate motion measure.

If a pixel has high value of motion measure then most probable it is a moving pixel. But if the motion measure is low, we can't necessarily predict it as stationary, as it may be uniform intensity region or region along an edge. To confirm stationarity of a pixel, we compute reliability of the motion measure. A pixel can be assumed stationary only if it has low motion measure along with high reliability.

Motion measure and reliability measure are defined as follows:

The *motion measure* is defined as "the weighted average of the normal flow magnitudes over a small (3x3 neighborhood) neighborhood N(x, y) of (x, y)".

$$MM = \frac{\sum_{(x_i,y_i\ )\in N(x,y)} |I(x_i, y_i, t+1) - I(x_i, y_i, t)|.|\nabla I(x_i, y_i, t)|}{\sum_{(x_i,y_i\ )\in N(x,y)} |\nabla I(x_i, y_i, t)|^2 + C}$$

Where, $\nabla I(x, y, t)$: Intensity gradient at time t

$\nabla I(x, y, t)^{\ 2}$ is taken as weight.

C = constant used for avoiding numerical instability.

The reliability measure of the motion is calculated based on the numerical stability of equation (8) reproduced here as-

$$\begin{bmatrix} \left(\sum_{(x,y)\in N} I_x\right)^2 & (\sum_{(x,y)\in N} I_x I_y) \\ (\sum_{(x,y)\in N} I_x I_y) & \left(\sum_{(x,y)\in N} I_y\right)^2 \end{bmatrix} \begin{bmatrix} t_x \\ t_y \end{bmatrix} = \begin{bmatrix} -\left(\sum_{(x,y)\in N} I_t I_x\right) \\ -\left(\sum_{(x,y)\in N} I_t I_y\right) \end{bmatrix}$$

Reliability measure is defined as:

$$RM = \frac{\lambda_{min}}{\lambda_{max}}$$

Where, λ_{min} : Smallest Eigen value of the coefficient matrix

λ_{max} : Largest Eigen value of the coefficient matrix

Motion measure and reliability measure are used to define motion filter. If a pixel has high value of motion measure, it is certainly a moving pixel. But if a pixel is having low value of motion measure, then its reliability measure is checked, if reliability measure of the pixel is high then the pixel is classified as stationary. If a pixel is having low motion measure with low reliability measure then the pixel is an unknown pixel.

As we have studied from previous research in this field and our results using only motion estimation based classification, motion information alone is not sufficient to provide good result. Hence along with motion information we have used background subtraction.

For classification of unknown pixel as moving or stationary, we have applied the background subtraction using earlier estimated background (background initialized with the first frame and updated during the process) and have taken the difference between the transformed $frame_{(t+1)}$ and background and compared it to a threshold. If the difference exceeds the threshold value, we classify the unknown pixel as moving else the pixel is stationary. Moving pixels are considered as foreground and stationary pixels as background for foreground and background labeling based on motion estimation.

Based on the above process we get the foreground mask (FGM) and background mask (BGM) for the current frame. We use these masks to generate the background for the current frame as

$$BG = BG_M \text{ X } I + FG_M \text{ X } BG$$

3.3.2 Background Subtraction

For final foreground and background labeling we have taken the difference between $frame_{(t+1)}$ and new background and the difference is compared to a threshold. If the difference exceeds the threshold value, then we consider it as foreground else background.

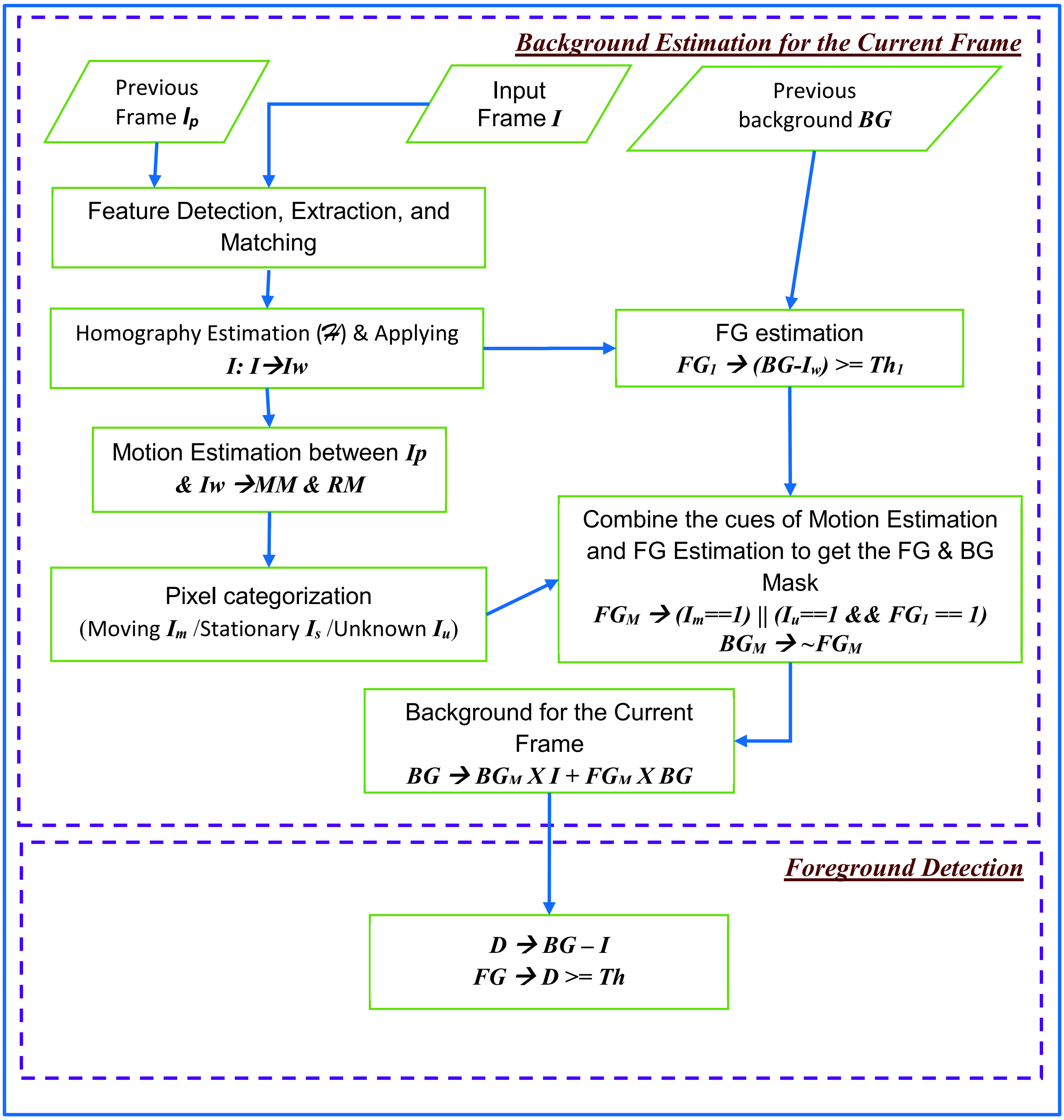


Figure 10: Proposed method block diagram

3.3.3 Algorithmic steps

1. Read frame t and (t+1). Consider first frame as background B.
2. Extract and match feature points using optical flow feature extraction.
3. Calculate homography matrix H using matching features and MSAC algorithm and exclude the outliers.
4. Transform the $(t+1)^{th}$ frame using this homography matrix H.
5. Calculate motion measure (MM):

$$MM = \frac{\sum_{(x_i,y_i)\in N(x,y)} |I(x_i, y_i, t+1) - I(x_i, y_i, t)|.|\nabla I(x_i, y_i, t)|}{\sum_{(x_i,y_i)\in N(x,y)} |\nabla I(x_i, y_i, t)|^2 + C}$$

$$Here\ |\nabla I(x_i, y_i, t)| = \left|\sqrt{(\frac{\partial I}{\partial x})^2 + (\frac{\partial I}{\partial y})^2}\right|$$

$$I(x_i, y_i, t+1) : Transformed\ frame\ \ using\ H$$

$$I(x_i, y_i, t) = Original\ frame$$

6. Calculate reliability measure (RM):

$$RM = \frac{\lambda_{min}}{\lambda_{max}}$$

7. Define motion filter as:
   If MM is high => moving pixel
   If MM is low and RM is high => stationary pixel
   If MM is low and RM is low => Unknown pixel
8. For unknown pixel classification calculate difference d = $(t+1)^{th}$ frame – B.
9. Set threshold T_1
10. If d > T_1, pixel is moving else stationary.
11. Moving pixels are foreground mask and stationary pixels are background mask.
12. Update B as new background:

$$New\ Background = (1 - Foreground)frame_{(t+1)} + Foreground.B$$

13. Take difference d = $(t+1)^{th}$ frame – new background.

14. Set threshold T_2

15. If $d > T_2$, foreground else background.

CHAPTER 4: EXPERIMENTAL RESULTS

Frame no.	Input Image	Bounding Box	Background Mask	Result/ Foreground
189				
194				
201				
211				
221				
231				

241				
251				
256				
261				

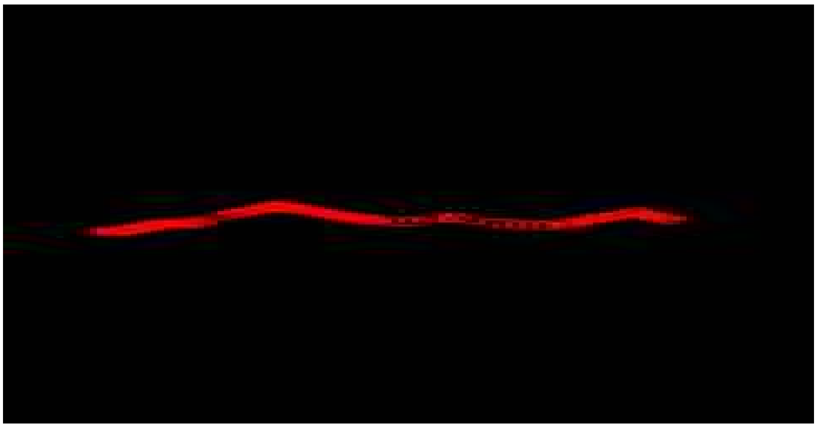


Figure 11: Tracked Path

Frame no.	Input Image	Bounding Box	Background mask	Result/ Foreground
75				
85				
95				
105				
115				
121				
125				

135				
145				
151				

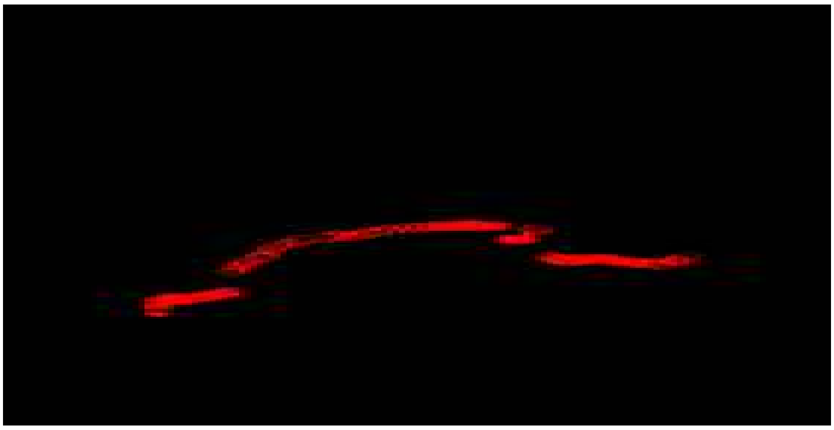
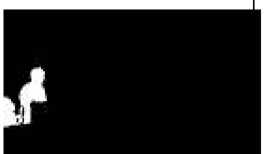
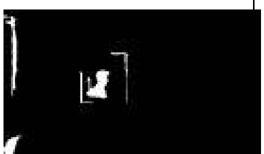
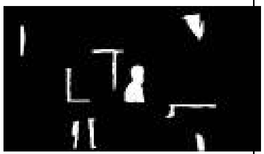
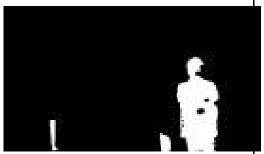
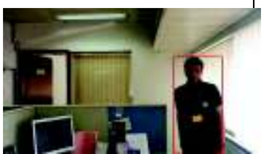


Figure 12: Tracked Path

Frame no.	Input Image	Bounding Box	Background Mask	Result/ Foreground
75				
76				
78				
82				
85				
88				
125				

143				
185				
189				
199				
200				

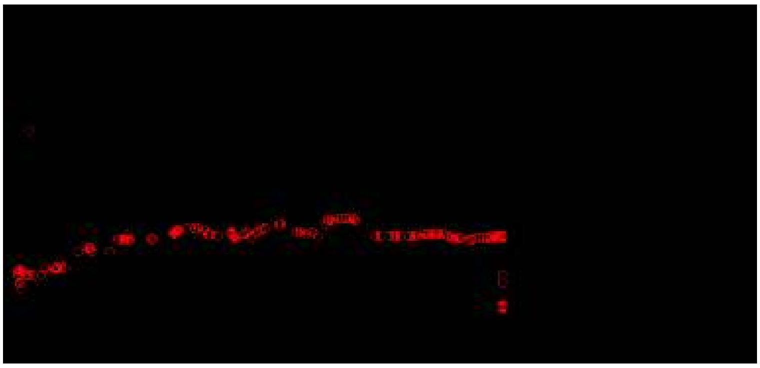


Figure 13: Tracked Path for Office Dataset

The proposed method is implemented in MATLAB R2014b. The result has been tested on many datasets for performance evaluation. Some of the results have been displayed here.

In the results, we have shown frame numbers of the input sequence, bounding box tracker, the modelled background and the final result in 1, 2, 3, 4, and 5 columns respectively. We have shown here results of three different datasets. By analyzing the above results, it can be concluded that the results are good enough and the method proposed for "Moving Foreground Object Detection and Tracking in Visual Surveillance System" is a significant contribution in this area of research.

CHAPTER 5: CONCLUSION AND FUTURE WORK

In this project, a new method for moving object tracking in moving camera scenario was proposed. It was achieved in two steps, background estimation and background subtraction. For background estimation, brightness constancy constraint of optical flow was considered and motion measure (MM) and reliability measure (RM) was calculated for each pixel. Based on this MM, RM, and the cue obtained from the difference of current frame and background, the pixel state was classified as moving or stationary and foreground and background masks was obtained. Using these masks and the previous background, new background was modelled for the current frame. And then background subtraction was performed to get foreground and background. Experiment was performed on several datasets and then the experimental results were investigated and it was concluded that the results was reasonably good and the proposed method outperforms most of the other existing methods.

The proposed method was able to track the object present even in the first frame of the video. This method also handled fast moving background in many cases. And the result were reasonably good even under fast moving background and was able to detect and track object. But the detection quality in these challenging cases was not so good. In future, we need to develop more robust methods for object detection and tracking, which could handle these challenging situations.

REFERENCES

[1] C. Chattopadhyay, S. Das, "Prominent moving object segmentation from moving camera video shots using iterative energy minimization", Int. Conf on SIViP, Springer, May 2014.

[2] A. Romanoni · M. Matteucci, D. G. Sorrenti, "Background subtraction by combining Temporal and Spatio-Temporal histograms in the presence of camera movement", Machine Vision and Applications, Springer, vol. 25, pp. 1573–1584, 2014.

[3] S. W. Kim, K. Yun, K. Y. Yi, S. J. Kim, J. Y. Choi, "Detection of moving objects with a moving camera using non-panoramic background model", Machine Vision and Applications, Springer, vol. 24, pp. 1015-1028, Aug 2012.

[4] S. L. Vaikole, S. D.Sawarkar, "Segmentation of Moving Object in Video with Camera in Motion", International Conference on Nascent Technologies in the Engineering Field (ICNTE-2015), IEEE, pp. 1-5, 2015.

[5] W. J. Kim, I. Kweon, "Moving Object Detection and Tracking from Moving Camera", 8th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), IEEE, Nov. 2011.

[6] A. Ghosh, B. N. Subudhi, S. Ghosh, "Object Detection From Videos Captured by Moving Camera by Fuzzy Edge Incorporated Markov Random Field and Local Histogram Matching", IEEE Transactions On Circuits And Systems for Video Technology, vol. 22, no. 8, pp. 1127-1135, Aug. 2012.

[7] Y. Jin, L. Tao, H. Di, N. I. Rao, G. Xu, "Background Modeling from a Free-Moving Camera by Multi-Layer Homography Algorithm", 15th International Conference on Image Processing, IEEE, pp. 1572-1577, aug. 2008.

[8] M. Irani M., B. Rousso, S. Peleg, “Computing occluding and transparent motions”, Int. Journal of Computer Vision, volume 12, issue 1, pp. 5-16, Feb. 1994.

[9] Cris Stauffer and W. E. L. Grimson, “Adaptive background mixture models for real-time tracking”, in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 246-252, June 23-25, 1999.

[10] Mahfuzul Haque, Manzur Murshed, and Manoranjan Paul, “On stable dynamic background generation technique using Gaussian mixture models for robust object detection”, in 5th International Conference on Advanced Video and Signal Based Surveillance, IEEE, pp. 41-48, September 1-3, 2008.

[11] C. R. Jung, “Efficient Background Subtraction and Shadow Removal for Monochromatic Video Sequences”, IEEE Trans. on Multimedia, Vol. 11, No. 3, pp. 571-577, 2009.

[12] Aparna Akula, Nidhi Khanna, Ripul Ghosh, Satish Kumar, Amitava Das, and H. K. Sardava, “Adaptive Contour based statistical background subtraction method for moving target detection in infrared video sequences”, Journal of Infrared Physics and Technology, Elsevier, vol. 63, page 103-109, Dec. 2013.

[13] A. Tewary, A. Akula,R. Ghosh, S. Kumar, H K Sardana, "Hybrid multi-resolution detection of moving targets in infrared imagery", Journal of Infrared Physics and Technology, Elsevier, vol. 67, pp. 173-183, July 2014.

[14] X. Dong, X. Huang, Y. Zheng, S. Bai, W. Xu, “A novel infrared small moving target detection method based on tracking interest points under complicated background”, Journal of Infrared Physics and Technology, Elsevier, vol. 65, page 36-42, April 2014.

[15] V. Reddy, C. Sanderson, B. C. Lovell, “Improved Foreground Detection via Block-based Classifier Cascade with Probabilistic Decision Integration”, IEEE Transactions On Circuits And Systems for Video Technology, vol. 23, no.1, Jan. 2013.

[16] X. Zhou, C. Yang, W. Yu, “Moving Object Detection by Detecting Contiguous Outliers in the Low-Rank Representation”, IEEE Trans. On Pattern Analysis and Machine Intelligence, volume 35, issue 3, pp. 597 – 610, 2013.

[17] T. S. F. Haines and T. Xiang, “Background Subtraction with Dirichlet process mixture model”, IEEE Transactions on Pattern Analysis Machine Intelligence, vol. 36, no. 4, pp. 670-683, 2014.

[18] J. Dou, J. Li, “Moving object detection based on improved VIBE and graph cut optimization”, ELSEVIER 2013.

[19] P. L. St-Charles, G. Bilodeau, and R. Bergevin, ‘SuBSENSE: A universal change detection method with local adaptive sensitivity’, IEEE Transactions on Image Processing, vol. 24, issue 1, pp. 359-373, January, 2015.

[20] W. Choi, C. Pantofaru, S. Savarese, "A General Framework for Tracking Multiple People from a Moving Camera", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 7, pp. 1577-1591, July 2013.

[21] Dileep Kumar Yadav, “Efficient method for moving object detection in cluttered background using Gaussian mixture model”, in 3rd International Conference on Advances in Computing, Communications and Informatics, IEEE, pp. 943-948, September 24-27, 2014.

[22] A. Yilmaz, O. Javed, and M. Shah. “Object Tracking: A Survey”, ACM Computer Survey 38, 4, Article 13, December 2006.

[23] D. K. Yadav, and K. Singh, “Motion based object detection in real-time visual surveillance system using adaptive learning”, Journal of Information Assurance and Security, vol. 10, issue 2, pp. 89-99, March, 2015.

[24] G. Bradski, and A. Kohler, Learning OpenCV, Computer Vision with OpenCV Library, O’Reilly Publication, 2011.

[25] S. Lee, C. Lee, “Low-Complexity Background Subtraction based on spatial similarity”, Eurasip Journal on Image and Video Processing, Springer, volume 1, issue 30, pp. 1-16, 2014.

[26] D. K. Yadav, L. S. Sharma, S. K. Bharti, “Moving Object Detection in Real-Time Visual Surveillance using Background Subtraction Technique” 14th International Conference on Hybrid Intelligent Systems, IEEE, pp. 79-84, December 14-16, 2014.

[27] H. Yao and I. S. Ahmad, "Adaptive Foreground Segmentation Using Fuzzy Approach", 4th International conference on digital information management, IEEE, pp. 103-108, November, 2009.

[28] Z. P Dordevic, S. G Graovac, and S. T Mitrovic, “Suboptimal threshold estimation for detection of points-like objects in radar images”, Eurasip Journal on Image and Video Processing, Springer Open Access Journal, vol. 5, pp. 1-12, February, 2015.

[29] L. Lin, Y. Xu, X. Liang, and J. Lai “Complex Background Subtraction by Pursuing Dynamic Spatio-Temporal Models”, IEEE Transactions on Image Processing, vol. 23, issue 7, pp. 3191-3202, July, 2014.

[30] P. Suo, and Y. Wang, "An improved adaptive background modeling algorithm based on Gaussian Mixture Model", in 9th International Conference on Signal Processing, IEEE, pp. 1436-1439, October 26-29, 2008.

[31] R. Zhang, W. Gong, V. Grzeda, A. Yaworski, and M. Greenspan, "An Adaptive Learning Rate Method for Improving Adaptability of Background Models", IEEE Signal Processing Letters, vol. 20, no. 12, pp. 1266-1269, December, 2013.

[32] JL Barron, DJ Fleet, SS Beauchemin, "Performance of Optical Flow Techniques", International Journal of Computer Vision, vol. 12, issue 1, pp. 43-77, 1994.

[33] D. Zamalieva, A. Yilmaz, J. W. Davis, “A Multi-transformational Model for Background Subtraction with Moving Cameras”, ECCV 2014, Part I, LNCS 8689, pp. 803–817, 2014.

[34] Ali Elqursh and Ahmed Elgammal, “Online Moving Camera Background Subtraction”, ECCV 2012.

[35] Y. Sheikh, O. Javed, T. Kanade, “ Background Subtraction for Freely Moving Cameras”, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, Object Video Inc., Reston, VA 20191.

[36] M. Narayana, A. Hanson, E.L.-Miller, “Coherent Motion Segmentation in Moving Camera Videos using Optical Flow Orientations”, University of Massachusetts, Amherst.