Object Detection and Tracking in Video Surveillance System In Low Resolution Condition

1 Neeraj Kumar Pandey, 2 Dr. Brij Mohan Singh,
1Assistant Professor, Deptt. Of CS, College of Engineering Roorkee, (Uttarakhand)-India.
2Associate Professor, Deptt. Of CS, College of Engineering Roorkee, (Uttarakhand)-India.

Abstract: This paper presents a detailed view of various techniques related to video surveillance system improving the security. The goal of this paper is to review the various moving object detection and object tracking methods. This paper focuses on detection of moving objects in video surveillance system then tracking the detected objects in the scene in all condition like low resolution. Moving Object detection is first low level important task for any video surveillance application. Detection of moving object is a challenging task. Tracking is required in higher level applications that require the location and shape of object in every frame. The conventional approach to object tracking is based on the difference between the current image and the background image. However, algorithms based on the difference image cannot simultaneously detect still objects. Furthermore, they cannot be applied to the case of a moving camera. Algorithms including the camera motion information have been proposed previously, but, they still contain problems in separating the information from the background. Video Surveillance Systems have been in use for long. It has several applications such as Security in public and commercial domains, Smart video data mining and Law enforcement and Military purposes etc. Video surveillance is all about keeping track of activities at a particular place through surveillance cameras. It uses the concept of Detecting and Tracking different objects in all condition like low resolution to detect any suspicious activity and take any kind of action required. There are also some other areas where Video Surveillance is used like Patrolling of highways and railways for accident detection, Surveillance of properties and forests for fire detection, Observation of the activities of elderly and infirm people for early alarm, Measuring effectiveness of medical treatments. In this paper, we propose an approach for object detection from a sample input video and then it performs object tracking using Kalman Filter.

Keywords: Object detection, tracking, low resolution

I. INTRODUCTION

Computer Vision is a field that includes methods for acquiring, processing, analysing and understanding images and in general, high dimensional data from real world in order to produce numerical or symbolic information, e.g., in form of decisions. Detecting objects in an image and tracking them through a video sequence has been a fundamental problem in computer vision and has led to a significant amount of research in the last few decades. One of the most important step in video surveillance is object detection and tracking. Object detection from an input video and tracking it or in other words, the creation of temporal correspondence among detected objects from frame to frame is a part of Computer Vision.

Videos are basically the sequence of images. Each image is called a frame, which is displayed in fast enough frequency so that human eyes can percept the continuity of the sequence. All image processing techniques can be applied to individual frames. In addition to this, the contents of two consecutive frames are closely related.

Object detection in videos involves verifying the presence of the object in the video or in other words, sequence of images and locating it for recognition. Object tracking is done to monitor
Object’s spatial and temporal changes during a video sequence, which includes its presence, position, size, shape, etc. It is done by solving the temporal correspondence problem or the problem of matching the target region in successive frames of a video taken at close time intervals. These two processes are closely related because tracking generally begins with detecting objects, while detecting an object repeatedly in an image sequence is often necessary to help and verify tracking.

Tracking detected objects frame by frame in video is a significant and difficult task. It is a crucial part of smart surveillance systems since without object tracking, the system could not extract cohesive temporal information about objects and higher level behaviour analysis steps would not be possible. Moving object detection is the first step in video analysis. It can be used in many regions such as video surveillance, traffic monitoring and people tracking.

The objective of tracking is to establish correspondence of objects and object parts between consecutive frames of video. It is a significant task in most of the surveillance applications since it provides cohesive temporal data about moving objects which are used both to enhance lower level processing such as motion segmentation and to enable higher level data extraction such as activity analysis and behaviour recognition.

Object Detection and Tracking is all about taking an input video, extracting a sequence of frames from it, detecting and tracking the desired object and analysing its activity.

This paper is basically a video surveillance system which is built for providing security in many forms. Adding an additional feature of Intelligence to this system provides us with a concept of Motion Detection without Human intervention which is particularly useful in highly secured areas. Our paper goal is to construct discriminative underlying human action representations, make computer beware of human action and recognize human behaviours in various scenarios effectively and efficiently.

If the activity detected is found to be suspicious i.e., matching with any kind of suspicious activity in our database, an alarm is to be raised to aware everyone about it or any different possible action needed to be taken.

This paper is proposed for the Detection and Tracking of Object using Computer vision based approach. The objective of this paper is to achieve good results for Object Detection, Tracking and Activity Recognition. This paper aims at extracting frames from an input video such as an AVI, FLV, MPEG file etc. and then detecting and tracking the movement of the desired object for Action Recognition.

Analysing the actions of humans by using cameras can be termed as Action Recognition. This concept of Action Recognition is now used in many fields, especially in the field of Robotics and intelligent systems in which there is a greater need for the recognition of the actions. Recognizing of humans and also their activity is very much important for any intelligent system, which is to be done intelligently and effortlessly with a human-inhabited environment.

It is a challenging research area because the dynamic human body motions have unlimited underlying representations and to recognize each of the behaviour correctly one needs to build a large and efficient database by deeply analysing each activity of human behaviour.

The proposed method extracts low-level features from a video stream during a short time period for describing human actions, and the extracted sequence of the low-level features is used for describing human activity.

II. LITERATURE REVIEW

The importance and popularity of object detection and tracking has led to several previous works. A review of all the past work done in this field is presented here:
In year 2004, Zaveri et al. [1] proposed an algorithm for detection and tracking of small and fast moving objects, like a ping pong ball or a cricket ball, in sports video sequences. For detection, the proposed method uses only motion as a cue; moreover it does not use any texture information. The method is able to detect the object with very low contrast and negligible texture content. Along with detection, they also proposed a tracking algorithm using the multiple filter bank approach. Thus they provide a complete solution. The tracking algorithm is able to track manoeuvring as well as non-manoeuvring movements of the object without using any a priori information about the target dynamics.

In year 2004, Mei Han et al. [2] described a method for tracking multiple objects whose number is unknown and varies during tracking. Based on preliminary results of object detection in each image which may have missing and/or false detection, the multiple object tracking method keeps a graph structure where it maintains multiple hypotheses about the number and the trajectories of the objects in the video. The image information drives the process of extending and pruning the graph, and determines the best hypothesis to explain the video. While the image-based object detection makes a local decision, the tracking process confirms and validates the detection through time; therefore, it can be regarded as temporal detection which makes a global decision across time. The multiple object tracking method gives feedbacks which are predictions of object locations to the object detection module. Therefore, the method integrates object detection and tracking tightly. The most possible hypothesis provides the multiple object tracking result. The experimental results are presented.

In year 2008, Huang, T.S.; et al. [3] presented an effective method small object localization that integrates detection and tracking. The system is initialized using a strong detector and then it locates the object over time using a weak detector and a temporal tracker. Both of strong and weak detectors are based on foreground-background segmentation. The strong detector is created from shape analysis of foreground blobs and used to trigger the object tracker. The weak detector is built with outputs from the foreground detection likelihood and integrated into the tracker observation likelihood. In the particle filter-based object tracker, motion estimation is embedded to generate a better proposal distribution and a mixture model is tailored to handle the ambiguity of template matching due to cluttered background. As a case study, the proposed scheme is applied to ball detection and tracking in soccer game videos. Promising results are presented to illustrate the proposed method effectively handles heavy clutter, occlusion and motion blur.

In year 2008, Leibe, B. [4] presented a novel approach for multi-object tracking which considers object detection and spacetime trajectory estimation as a coupled optimization problem. The approach is formulated in a minimum description length hypothesis selection framework, which allows our system to recover from mismatches and temporarily lost tracks. Building upon a state-of-the-art object detector, it performs multiview/multicategory object recognition to detect cars and pedestrians in the input images. The 2D object detections are checked for their consistency with (automatically estimated) scene geometry and are converted to 3D observations which are accumulated in a world coordinate frame. A subsequent trajectory estimation module analyses the resulting 3D observations to find physically plausible space-time trajectories. Tracking is achieved by performing model selection after every frame. At each time instant, our approach searches for the globally optimal set of space-time trajectories which provides the best explanation for the current image and for all evidence collected so far while satisfying the constraints that no two objects may occupy the same physical space nor explain the same image pixels at any point in time. Successful trajectory hypotheses are then fed back to guide object detection in future frames. The optimization procedure is kept efficient through incremental computation and conservative hypothesis pruning. We evaluate our approach on several challenging video sequences and demonstrate its performance.
on both a surveillance-type scenario and a scenario where the input videos are taken from inside a moving vehicle passing through crowded city areas.

In year 2009, Cheng-Ming Huang et al. [5] presented a real-time tracking system to detect and track multiple moving objects on a controlled pan-tilt camera platform. In order to describe the relationship between the targets and camera in this tracking system, the input/output hidden Markov model (HMM) is applied here in the well-defined spherical camera coordinate. Since the detection and tracking for different targets are performed at the same time on a moving camera platform, the detection and tracking processes must be fast and effective. A hybrid detection algorithm which combines the target's colour and optical flow information is proposed here. A two layer tracking architecture is then utilized for tracking the detected target. The bottom level utilizes the Kanade-Lucas-Tomasi (KLT) feature point tracker which identifies the local point correspondence across image frames. The particle filter at top level, which maintains the relation between target and feature points, estimates the tracked target state. The overall performance has been validated in the experiments.

In year 2009, Ali, S.S.; et al. [6] proposed a robust moving object detection method in videos and discusses its applications to human and vehicle detection. Our method consists of average background model with supportive secondary model and an adaptive threshold selection model based on Gaussian distribution. The average background model is used for background modelling as used in [Narayana, 2007] and the background subtraction system is used to provide foreground image through difference image between current image and model image. The adaptive threshold method is used to simultaneously update the system to environment changes. This method is tested on various environments and experimental results show that proposed method is more robust and efficient than others in video-based object detection and tracking.

In year 2010, Liang Xiao et al. [7] proposed an approach to detect and track objects, some approaches of detecting and tracking moving objects in stationary scene are presented, including the detection methods of the time domain differential method, the background differential method, optical flow method and tracking methods of the Kalman filter and feature optical flow.

In year 2010, Hao Sun et al. [8] presented approach for detecting and tracking independently moving objects from a mobile platform using uncalibrated stereo cameras. Firstly, scale invariant feature transform (SIFT) features are detected and a novel multiview matching method is proposed for simultaneous stereo matching and motion tracking of the detected features. A multiview geometric constraint, derived from the relative camera positions in pairs of consecutive stereo views, is then derived for independent motion detection. Finally, a dimensional variable particle filter is introduced for joint detection and tracking of multiple independently moving objects. Experimental results on real-world stereo sequences demonstrate the effectiveness and robustness of our method.

In year 2013, Naeem, H et al. [9] discussed some famous and basic methods of object detection and tracking. Real-time object detection and tracking is a vast, vibrant yet inconclusive and complex area of computer vision. Due to its increased utilization in surveillance, tracking system used in security and many others applications have propelled researchers to continuously devise more efficient and competitive algorithms. However, problems emerge in implementing object detection and tracking in real-time; such as tracking under dynamic environment, expensive computation to fit the real-time performance, or multi-camera multi-objects tracking make this task strenuously difficult. In the end they have also given their general applications and results.

In year 2015, Shuai Zhang et al. [10] discussed new object detection, tracking, and recognition approaches for video surveillance over camera network. A new object detection algorithm using mean shift(MS) segmentation is introduced, and occluded objects are further separated with the help of depth information derived from stereovision. The detected objects are then tracked by a
new object tracking algorithm using an overlapped Bayesian Kalman filter with simplified Gaussian mixture (BKF-SGM).

From all the above research work and reviews carried out earlier, there are several methods being discovered to perform object detection and tracking effectively and efficiently to give better results. This paper is intended to perform object detection and tracking in a dynamic background i.e. where the background objects are not stationary. The process of Object Detection will be done using Feature Extraction and Feature Matching and can also be done using Gaussian Mixture Models using Blob Analysis. The process of Object Tracking will be performed using Kalman Filter or by Extracting Histograms of Oriented Gradients (HOG) features. The task of doing Activity Analysis or Action Recognition is research based which is intended to be listed as a Future Work in this paper.

III. OBJECT DETECTION AND TRACKING

As detection phase represent each object in the form of a point over the spatial domain the tracking is the process to assign consistent label to each detection point across the frame. Numerous approach for object tracking has been proposed that are depend upon object representation, image features and motion, appearance and shape of the object Detecting objects in an image and tracking them through a video sequence has been a fundamental problem in computer vision and has led to a significant amount of research in the last few decades.

In this paper, we propose an approach for object detection from a sample input video and then it performs object tracking using Kalman Filter or by extracting Histogram of Gradient Features.

IV. OBJECT DETECTION

Object Detection is an important phase where we have to detect the desired human from the input video.

At first, we divide the input video into a sequence of frames and then apply any of these algorithms for extracting an object which can be either human or non-human from a static background:

a) Background Subtraction algorithm
b) Temporal Differencing
c) Optical Flow

a) Background Subtraction algorithm

Background subtraction, also known as Foreground Detection, is a technique in the fields of image processing and computer vision wherein an image’s foreground is extracted for further processing (object recognition etc.).
Generally an image's regions of interest are objects (humans, cars, text etc.) in its foreground. After the stage of image pre-processing (which may include image denoising etc.) object localisation is required which may make use of this technique.

Background subtraction is a widely used approach for detecting moving objects in videos from static cameras. The rationale in the approach is that of detecting the moving objects from the difference between the current frame and a reference frame, often called “background image”, or “background model”. Background subtraction is mostly done if the image in question is a part of a video stream.

There are many constraints in developing a good background subtraction algorithm:

1. First, it must be robust against changes in illumination.
2. Second, it should avoid detecting non-stationary background objects such as moving leaves, rain, cast by moving objects.
3. Lastly, its internal background model should react quickly to changes in background such as starting and stopping of vehicles.

b) **Temporal Differencing algorithm:** Temporal differencing method uses the pixel-wise difference between two or three consecutive frames in video imagery to extract moving regions. It is a highly adaptive approach to dynamic scene changes however, it fails to extract all relevant pixels of a foreground object especially when the object has uniform texture or moves slowly. When a foreground object stops moving, temporal differencing method fails in detecting a change between consecutive frames and loses the object.

This method is computationally less complex and adaptive to dynamic changes in the video frames. In temporal difference technique, extraction of moving pixel is simple and fast. Temporal difference may left holes in foreground objects, and is more sensitive to the threshold value when determining the changes within difference of consecutive video frames. Temporal difference require special supportive algorithm to detect stopped objects.

Unlike, the background subtraction method where the base image was the background image, here in temporal differencing the reference image is the previous images. Hence the previous frame is subtracted from the current image and the subtraction value must be greater than a threshold value in order to give a difference image.

c) **Optical flow algorithm:** Optical flow can be used to detect independently moving targets in the presence of camera motion; however most optical flow computation methods are very complex and are inapplicable to real-time algorithms without specialized hardware. Optical flow can be used to segment a moving object from its background provided the velocity of the object is distinguishable from that of the background, and has expected characteristics. Optical flow is the amount of image movement within a given time period.

d) **Illumination Processing:** After the detection of object we apply Illumination Processing for removing the distortion caused in the extracted image due to illumination.

e) **Removing Noise:** The next phase is to remove noise from the extracted image by applying suitable filters. Noise occurs in the image generally during Image Acquisition and the transmission of image.

Images taken with both digital cameras and conventional film cameras will pick up noise from a variety of sources. Many further uses of these images require that the noise will be (partially) removed - for aesthetic purposes as in artistic work or marketing, or for practical purposes such as computer vision.
V. OBJECT TRACKING

Object tracking is the process of locating a moving object (or multiple objects) over time using a camera. It has a variety of uses, some of which are: human-computer interaction, security and surveillance, video communication and compression, augmented reality, traffic control, medical imaging and video editing. Video tracking can be a time consuming process due to the amount of data that is contained in video. Adding further to the complexity is the possible need to use object recognition techniques for tracking, a challenging problem in its own right.

The objective of video tracking is to associate target objects in consecutive video frames. The association can be especially difficult when the objects are moving fast relative to the frame rate. Another situation that increases the complexity of the problem is when the tracked object changes orientation over time. For these situations video tracking systems usually employ an motion model which describes how the image of the target might change for different possible motions of the object.

To perform video tracking an algorithm analyses sequential video frames and outputs the movement of targets between the frames. There are a variety of algorithms, each having strengths and weaknesses. Considering the intended use is important when choosing which algorithm to use. There are two major components of a visual tracking system: target representation and localization, as well as filtering and data association.

Target representation and localization is mostly a bottom-up process. These methods give a variety of tools for identifying the moving object. Locating and tracking the target object successfully is dependent on the algorithm. For example, using blob tracking is useful for identifying human movement because a person's profile changes dynamically. Typically the computational complexity for these algorithms is low. The following are some common target representation and localization algorithms:

a) **Blob tracking**: segmentation of object interior (for example blob detection, block-based correlation or optical flow)
b) **Kernel-based tracking** (mean-shift tracking): an iterative localization procedure based on the maximization of a similarity measure (Bhattacharyya coefficient).
c) **Contour tracking**: detection of object boundary (e.g. active contours or Condensation algorithm)
d) **Visual feature matching**: registration

a) **Region Based Tracking**

Here, the features of the blob, detected in one image frame are matched to the blob detected in the other frame. If there is a match then the detected image is linked with the image in the previous frame.
Here, the energy of the boundary/contour of the blob detected in the previous frame is matched with the energy of the boundary of the blob detected in the current image.

In case of stationary background, region tracking is very efficient with stationary background. Hence we are taking region tracking. Now region tracking stores the features of whole object and matches the features with the features of the object in the next frame.

This wastes time, instead most salient region method only matches the most salient region of the previous frame with the most salient region of the current frame. Hence, reducing the amount of time required to match the whole image. The most salient region tracking works as follows:

The initial features for colour, orientation and intensity are fetched from the image.

From the above fetched features, the feature vector is calculated for colour, orientation and intensity using centre-surround method.

Once the feature map is obtained, the 3 features are weighted in order to find out which feature more uniquely identifies the object. Once weight is obtained, using their weight we get the saliency map for the detected object.

Finally feature weight vector is calculated for this most salient region. This feature weight vector is matched with the subsequent frame’s feature weight vector, if the match is above the threshold value then there is a match otherwise, the search area is doubled and the above procedure is repeated again.

Even after doubling the search area, a match is not found then object is expected to be occluded by some other object or any stationary background object.

Now, occlusion handling is not a separate process, it is a part of human tracking process. When an object is detected its centroid can easily be obtained and when a match is found between the object in the previous and the current frame object, we will get the centroid of the current image. From the centroids of these two images we will obtain following:

In case, the object is not found even after doubling the search area, the object can be said to be occluded and its next position can be predicated by the previously calculated velocity and direction.
of motion. In order to calculate the velocity and the direction of motion, minimum 5 frames should be processed.

VI. FILTERING AND DATA ASSOCIATION
Filtering and data association is mostly a top-down process, which involves incorporating prior information about the scene or object, dealing with object dynamics, and evaluation of different hypotheses. These methods allow the tracking of complex objects along with more complex object interaction like tracking objects moving behind obstructions. Additionally the complexity is increased if the video tracker (also named TV tracker or target tracker) is not mounted on rigid foundation (on-shore) but on a moving ship (off-shore), where typically an inertial measurement system is used to pre-stabilize the video tracker to reduce the required dynamics and bandwidth of the camera system. The computational complexity for these algorithms is usually much higher.

I used the following description of Kalman filter by Welch & Bishop [11][12]. The equations for the Kalman filter fall into two groups: time update equations and measurement update equations.

The time update equations: The time update equations are responsible for projecting forward (in time) the current state and error covariance estimates to obtain the priori estimates for the next time step.

The measurement update equations: The measurement update equations are responsible for the feedback — i.e. for incorporating a new measurement into the priori estimate to obtain an improved a posteriori estimate.

The first task during the measurement update is to compute the Kalman gain. After each time and measurement update pair, the process is repeated with the previous a posteriori estimate used to project or predict the new priori estimates. This recursive nature is one of the very appealing features of the Kalman filter - it makes practical implementations much more feasible than (for example) an implementation of a Wiener filter, which is designed to operate on all of the data directly for each estimate. The Kalman filter instead recursively conditions the current estimate on all of the past measurements.

![Figure: A complete picture of the operation of the Kalman filter](image)

Depending on the application, one might want to obtain an estimate of the state at a certain time. If the state is estimated for some future time, the process is called prediction. If the estimate is made using all measurements up to and including the current moment, one speaks of filtering. If an estimate is made for some time in the past using measurements until the current moment, the process is called smoothing [13]. In practice, the use of Kalman filter is limited by the ubiquitous nonlinearity and non-Gaussianity of physical world.
VII. ALGORITHM

STEP 1:
Initialization of webcam or video

STEP 2:
Getting the frame and then capturing the current frame and estimating the background.

STEP 3:
Subtracting the background frame from current frame
   \[ R = \text{current-back}; \]

STEP 4:
IF \( R \neq 0 \);
No object detected;
END;
Else if \( R = 0 \);
Region props-fixed bounding box (area, centroid etc)
Annotated(rectangle)
END;

VIII. EXPERIMENTAL RESULTS AND DISCUSSION

The Experimental Results for Object Detection and Tracking are explained in a stepwise manner below:

Step 1: Taking the input video and extracting frames from it. The Key Frames of the input video or recorded live via web cam

Step 2: Apply the process of Background Subtraction to each of the frames. After that, Image averaging is done to reduce the No. Of frames or to optimize them.

Step 3: Noise such as Salt and Pepper Noise and Gaussian Noise is removed from the optimized frames using Median Filter and Averaging Filter.

Step 4: Object Detection is performed using Feature Extraction and then using Feature Matching.
Step 5: Object Tracking is performed either by using Kalman Filter

IX. CONCLUSIONS AND FUTURE WORK

In this paper we have proposed a method for Object Detection, Tracking and Activity Analysis. Background subtraction method is used for segmentation of foreground object from background and to obtain maximum number of moving pixels. With Illumination estimation, the background model can be easily updated with change in luminance intensity by the determination of background.

Object Detection is performed using the optimized no. of Frames by performing Feature Extraction and then Feature Matching. Object Tracking is done using Kalman Filter or by Extracting Histograms of Oriented Gradients (HOG) features.

The algorithm efficiently tracks the moving object whose motion and appearance change drastically.

The future work is intended to working towards Activity Analysis which may help in detection of any suspicious activity at any crowded place effectively in order to trigger an alarm and ensure safety.

REFERENCES

2. Mei Han ; NEC Lab., Cupertino, CA, USA ; Sethi, A. ; Wei Hua ; Yihong Gong, “A detection-based multiple object tracking method”, 2004, IEEE
5. Cheng-Ming Huang; Dept. of Electr. Eng., Nat. Taiwan Univ., Taipei, Taiwan ; Yi-Ru Chen ; Li-Chen Fu, “Real-time object detection and tracking on a moving camera platform”, 2009, IEEE
12. Greg Welch and Gary Bishop “An Introduction to the Kalman Filter” University ofNorthCarolinaatChapelHill, Departmentof Computer Science Chapel Hill, NC 27599-3175, SIGGRAPH 2001 Course8