A Survey On Vehicle Detection And Tracking Algorithms In Real Time Video Surveillance

Sri Jamiya S, Esther Rani P

Abstract: Automated surveillance systems are of critical importance in traffic management and monitoring any unwanted activities. The intelligent transportation system plays a crucial role in the field of traffic management to provide an efficient and reliable transportation system. One of the applications of the intelligent transportation system is to detect and track vehicles accurately. Image processing algorithms has been widely developed to monitor the motion of vehicles, humans or any other objects. The main aim is to detect and recognize moving objects from real surveillance videos to avoid congestion on highways and parking areas for the prevention of accidents. In comparison with still images, each and every video frame provide intelligent information about vehicles in various scenarios that change over time. Many algorithms have been developed to improve efficient real-time detection of incidents and it is a challenging task for the researchers to determine the driver’s behavior even in the diversity of vehicles, weather and light conditions. In this paper, the detailed overview of object motion detection, classification, and tracking algorithms are presented and also their strengths and weakness of the various algorithms are discussed.

Index Terms: Intelligent Transportation System, Vehicle detection, Foreground detection, Object classification, Tracking, Feature Extraction, Occlusion, Surveillance Systems

1. INTRODUCTION

The technological advancement of intelligent transportation systems has been applied in developed countries to monitor traffic movement and for accurate tracking of moving vehicles by the use of CCTV cameras. In literature, a wide range of algorithms have been proposed for the implementation of intelligent transport systems. In this paper the various algorithms, their merits, and demerits are discussed in detail. Video surveillance systems are also widely used to monitor sensitive areas from security point of view for a very long time. They include first generation (1GSS), second generation (2GSS), and third generation (3GSS). The first generation (1GSS, 1960-1980) surveillance systems were generally based on analog subsystems for the acquisition of images and transmission. The videos of various monitoring cameras are sent to a central control room. The major drawbacks in this system is high bandwidth requirement, difficult to store data and retrieval of videos due to a large number of videotape requirements and difficult to monitor live videos as it requires human operators with least time span. The second-generation (2GSS, 1980-2000) surveillance systems were of hybrid type. They used both analog and digital subsystems to resolve drawbacks from first generations. They made use of the advantages of digital video processing to monitor real-time events by providing assistance to human operators. Third generation (3GSS, 2000- ) surveillance systems method defines end-to-end digital communication systems. Image acquisition and processing were assisted by sensors, communications have been done through mobile and data was stored under central servers which uses low cost digital infrastructure. In 3GSS the CCTV surveillance cameras are used to identify the apparent movement of objects, primarily for the purpose of public safety, crime prevention, accidents in roadside and highways. The visual surveillance systems can generate real-time alarms by the use of various techniques even in the complex environment like day/dark changes, occluded conditions like tree shadows, wind blowing, etc. The smart surveillance systems are used for object detection and tracking in motion which makes the system more dependable and more robust. An efficient algorithm provides the apparent movement of object detection, classification, and tracking in various scenarios, to obtain accurate information.

The object detection in motion is the foremost step in analyzing the video frames where it segments the moving objects from the static background objects. The most popularly used techniques for object detection are background subtraction, statistical models, frame differencing, temporal differencing and optical flow models. The smart surveillance systems provide robust and precise detection of moving objects, even in the presence of various environmental conditions like light changes, occluded objects like trees, shadows, etc. The second step is Object classification which classifies into pedestrian behavior, the flow of vehicles, animals, and so on. The moving object classification techniques can be categorized into two types such as shape based and motion based methods. Shape based methods define spatial information like the width/height coefficient, binary edges map and vehicle’s outer contour can be specified. In motion based methods, temporal information are determined where the median is calculated for each and every pixel to classify the vehicles like small sized, medium sized and over-sized vehicles. By the use of smart surveillance systems the natural phenomenon such as fire and smoke can also be sorted. The third step in video processing is object tracking, which is capable of tracking moving objects frame by frame simply defined as the creation of temporal correspondence among detected objects from frame to frame. The tracked outputs of moving vehicles are used to construct an efficient transportation system that segments, classifies the moving vehicles. The final step of visual surveillance systems is object recognition that recognizes and differentiates the moving objects for decision making purpose. The visual surveillance system in computer vision provides very efficient and accurate in vehicle detection and tracking to detect and track the autonomous vehicles for the safety of the public. The various algorithms are used for providing the operators with exact information of high-level data to make correct decisions.
and store the information for future verification. The progression of technology in video surveillance systems, and safer driver assistance systems provides researchers to understand more effectively and some scenarios and applications are described below.

1.1 A SURVEY IN VIDEO SURVEILLANCE
This survey discusses many research works on objects classification, detection, and tracking. Such a system is required for preventing crime and accidents to ensure the safety of the public.

2 MOVING OBJECT DETECTION
Each application that benefits from smart video processing have different needs, thus requires different handling objects. However, they hold something in common like moving objects. In each and every vision systems, detecting moving objects are common such as people and vehicles in the video. Moving object detection steps consists of preprocessing, feature extraction, classification, detection, and tracking.

The algorithms like background subtraction techniques, optical flow methods, statistical methods, frame differencing, temporal differencing are the subsequently used techniques which are described below.

2.1 BACKGROUND SUBTRACTION
The background subtraction technique is widely used for motion segmentation in many applications. It finds the moving regions in images by subtracting the initial image of pixels from a referenced background image which is formed by averaging images. If the subtracted pixel value is greater than the threshold then it is defined as foreground. To enhance detected regions post-processing operations like dilation, erosion, and closing are performed to reduce the noise level. Many approaches for background subtraction technique are performed in terms of foreground detection, background maintenance, and post-processing. Heikkila and Silven used the simplest version where $l_t$ is marked as foreground by a pixel at location $(x,y)$ in the current image

$$l_t(x,y) - B_t(x,y) = \tau$$

(1)

and the predefined threshold is $\tau$ [1]. The (IIR) Infinite Impulse Response filter was used to update the $B_t$ background image,

$$B_{t+1} = \alpha l_t + (1-\alpha)B_t$$

(2)

By eliminating small-sized regions and morphological closing was used to create foreground pixel maps. Though background subtraction techniques are effective, they lack in performance with dynamic changes such as stationary objects uncover the backgrounds (e.g. a parked bus moves out of the parking) or sudden light changes.

2.2 STATISTICAL METHODS
The statistical models has evolved to solve the limitations of fundamental background subtraction techniques. In the statistical method, the characteristic of individual pixel or group of pixels are considered to construct the background frame and statistics of background can automatically update during processing. This technique provides more reliable and effective in several scenarios like illumination changes, distortion caused by low resolution, roadside trees, and shadows. In this system, pixels are represented by its intensity values $(M)$ minimum and $(N)$ maximum intensity values and $(D)$ maximum intensity difference between any succeeding frames noticed at the initial stage where the scene has stationary objects. The pixel is considered as foreground in the current image if it satisfies the condition below.

$$|M(x,y) - l_t(x,y)| > D(x,y) \text{ or } |N(x,y) - l_t(x,y)| > D(x,y)$$

(3)

After thresholding, the detected foreground pixel contains noise which is removed by morphological erosion. In the eroded areas, the size of pixels was small, to fix this a series of erosion and dilation was performed on the foreground pixel map. Small-sized noisy pixel regions are also taken away by applying connected component labeling. The static area of the current image was replaced by new images to perform an operation with the statistics of the background pixels. In another technique, Stauffer and Grimson [2] presented an adaptive background mixture model for tracking in real time. They used online approximations to update the modeled mixture of Gaussians converted from pixels.
2.3 Temporal Differencing
In temporal differencing, the reference image is the previous image. A new image is obtained when the difference between the previous frame and the current frame is greater than the threshold value. This method is highly recommended in dynamic scene environments. But it has some disadvantages in the detection of moving objects. For eg., If an object is in single color it fails to detect the whole pixel regions even when it is moving. It is ineffective to static scenes. For high-level processing and to detect stopped static objects other techniques must be let in. Lipton et al., [3] presented two frames differencing, the foreground frame is defined by the following equation.

\[ |l(x,y) - l_{t-1}(x,y)| > \tau \]  

In order to resolve the defects in two frame differencing Collins et al., [4] used a hybrid three frames differencing method. Video surveillance and monitoring (VSAM) is the highly recommended technique for observing moving objects in the sequence of images. This hybrid algorithm uses motion segmentation by combining background subtraction with a three-frame differencing technique. This algorithm detects moving objects quickly.

2.4 Optical Flow
Optical flow is based on motion segmentation and it detects moving objects, even when the camera is moving. However, it is computationally complex and has more noise.

2.5 Shadow and Light Change Detection
The motion detection algorithms described above are used for real-time surveillance for years, which performs well in all environments such as indoor and outdoor. However, without special work, most of these algorithms are vulnerable to both local (e.g. shadows and highlights) and global illumination changes (e.g. The sun being covered/uncovered with clouds). Motion detection is inaccurate when the moving objects are followed by shadows. Object classification also fails in the presence of shadows and sudden light changes. In background subtraction and shadow detection method, pixels are represented by a color model that separates brightness from the chromaticity component. The pixels in the image are divided into four types (background, shaded background or shadow, high-lighted background and moving foreground object) by calculating the distortion of brightness and chromaticity between the scene and the current image pixels. The two methods by which the shadows can be detected are statistic and video based method. In statistic based method the intensity values in shadow region are less than that in background is analyzed. In video based method difference between neighboring pixels intensity, geometry, color, and brightness are calculated.

3 Object Classification
The different objects detected in the video contains moving regions such as vehicles, humans, animals, and so on. To track it without difficulty we need to distinguish the types of objects in the detected video to analyze it properly.

\[ \text{Dispersedness} = \frac{\text{Perimeter}^2}{\text{Area}} \]  

3.1 Motion Based Classification
Motion based classification usually distinguishes non-rigid objects (e.g. human) from rigid objects (e.g. vehicles) by temporal motion features. The proposed method based on moving object temporal self-similarity. If an object shows periodic motion, it's self-similarity measure also shows a periodic motion. The method depends on this clue to categorize moving objects by periodicity. Objects like rigid and non-rigid are also identified by Optical flow analysis. A. J. Lipton., [3] proposed a method which uses local optical flow technique in regions of detected objects. High residual flow is present in non-rigid objects (humans) whereas rigid objects (vehicles) shows little residual flow. The motion of pedestrians will have a periodicity due to the generation of residual flow. By using this, it can be differentiated from various objects such as vehicles.

3.2 Fire Detection
Discussions of fire detection in research papers are rare in the computer vision literature. Many methods exploit color and extract the motion of fire features. It also generates false alarms in the presence of fire-colored segments. The models like spectral, spatial, temporal are defined to detect a fire in the video. The spectral is defined as a pixel color probability density of fire. The spatial defines the spatial structure of the fire region and temporal is used to capture the spatial structure changes.

3.3 Object Detection
The most important technique in the field of intelligent transportation system is object detection. In this technique the targets such as cars and traffic signs are detected. For detection, the shape of the car and spatial and temporal information from traffic signs are the extracted features [5], [6]. Optical flow is a technique in motion segmentation and detection, it detects moving objects even when the camera is moving. This method detects and tracks moving objects in aerial views. Its accuracy is more than background subtraction method. However, optical flow methods are complex to process and has more noise. The result of the methods used by Horn and Schunck in [7] were promising to perform better than the methods used by Lucas and Kanade in [8] for aerial views of detecting frames in motion. A lot of research works are carried out based on Horn and Schunck and Lucas and Kanade methods for enhancing optical flow. Many algorithms are proposed to detect motion in different scenarios [9], [10], [11], [12], [13]. In [11] indoor fixed cameras are used by Optical flow for detection objects in video streams with existing brightness. It was applied in the software of motion detection for analyzing of motion level, motion region and the number of objects. The only drawback of this method is the change in object velocity and lights in the area. In static cameras, tracking of moving objects are done by combining motion segmentation and optical flow algorithm in [9]. Optical flow does not depend on foreground or background regions. It follows segmentation using pixel by pixel classification. In [14], Optical flow algorithm was used in the regions of silhouette 2-way ANOVA and brightness change was minimized by object segmentation. In [15] crowd monitoring in videos was done by Horn and Schunck. The proposed technique detects and tracks the outdoor scenes with based optical flow. In [12] the outdoor scenes are detected with edge detection and gradient based on optical flow. The edge detection based techniques are more robust and it is not vulnerable to light changes. For the cameras in motion, the moving things are detected by classification and motion clustering in [10]. In [9] Fusion Horn
and Schunck used small squares in images with aerial color for estimating flow field in the color plane and those fields are fused together. Lucas and Kanade, in [16] Optical flow method is combined with stereo camera for UAVs. In [17], [18] for urban area navigation a combined way of fusion control was used. Lucas and Kanade provide promising results if the flow is constant with the pixels neighborhood to them in [19],[20]. The equation was derived by the local neighborhood of least squares in [17], [21]. The Horn and Schunck technique gives the best results if the flow is smooth throughout the entire frames, e.g. for some neighborhood, the objects motion is not having any restrictions in [22], [23]. Most of the researchers uses hybrid methods for motion detection; the hybrid method uses two or more methods of different kinds to get rid of motion detection problems. In [17] two techniques of optical flow algorithms were compared and its performance was evaluated. In [23] eight methods of optical flow algorithms was tested on synthetically generated data with added noise and of high complexity. The author claims that the performance of method in [8] provided the best result. The area of moving object was detected by a hybrid method of temporal difference and optical flow in [24]. The difference between frames is calculated by the temporal difference method and the differential image is filtered using low pass filter and edge detection techniques. The optical flow algorithm is used to find the velocity from the spatiotemporal derivative of image intensity. In [24] for a static camera the results from temporal difference and optical flow technique were quite promising but not for the camera in motion. Many motion detection algorithms were proposed and most of them use simple operation of thresholding on the difference in image intensity like the initial frame is compared with background frames from consecutive frames of videos, which depends on algorithms of the simplest form, yet its performance is not promising in [25]. In [26], [27], [28], [29] statistical and probabilistic models are used to increase the performance by background subtraction. The performances of these algorithms mainly depend on threshold value. In [25] various methods of threshold adaptation are described. By choosing Markov Random Field (MRF) in the framework of Bayesian [29] most promising results of detection are found by frame differencing and modeling change labels. In [30], Bayesian Markov random field (MRF) method was used to increase the performance in the use of shape for the detected objects and noise reduction. However, the most important work is to remove the background and pixels correlation in the frames. Bayesian algorithm [29] was used to extract the shape of the moving object. It deals very well with duplicate motions of objects, variation in light, reducing noise and removing shadows, the results and performance of the algorithm are promising. In [31], sobel filtering method was used for low quality web cameras in laptops to process moving images while using the same low end hardware. The initial algorithm is quite fast and the next one deals with edge detection of the object. The result of the algorithm shows 45.5% time of object detection with 14% use of memory, maintaining the same level of accuracy. In [32] Enhanced Dynamic Bayesian Network (DBN) technique is used for vehicle detection in surveillance, in the aeronautical field and this method is found to be flexible. In [33] multiple objects are tracked using spatial and second derivative detection and tracking model. But, it cannot track multiple objects in low quality videos. The Speeded Up Robust Features (SURF) deals to optimize Scale Invariant Feature Transform (SIFT) [34]. But SURF processing time is too long. Oriented FAST and Rotated Binary Robust Independent Elementary Features (ORB) algorithm [6] feature extraction is processed in an outdoor environment. For binary descriptor, the method of Local Difference Binary (LDB) is used. The image descriptors are matched by K-nearest Neighbor (KNN) [35]. The local invariant features are extracted by BRIEF alias ORB in [36]. The BRIEF technique is fast but had a drawback to noise. The advantage of BRIEF is it deals with two major problems, ORB is used to detect the corners with the help of the Harris method and uses intensity centroid for calculating the rotation of object direction in [37]. The researchers of this paper developed an algorithm to calculate the details in real time traffic by classification, counting of vehicles and segmentation. The important goal of this technique is it can work with the sudden change in light conditions by using feature based counting technique in detection of the vehicle and its tracking [38].

<table>
<thead>
<tr>
<th>Authors</th>
<th>Approach</th>
<th>Description</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>[7]</td>
<td>Horn and Schunck</td>
<td>Optical flow detects moving objects even when the camera is also in motion. It deals with the pattern of lights in the image for detection.</td>
<td>It can deal with a sequence of images which can be classified as a set rather than unshaped regions in spatial arrangements.</td>
<td>It is insensitive to noise and brightness levels.</td>
</tr>
</tbody>
</table>
The Classification methods proposed in this chapter are based on non-probabilistic algorithms which provide better results and are easy to implement [39], [40]. Some of these algorithms are given below: Fisher’s Linear Discriminant algorithm consists of two classes (-1,1), when the classification is made for an object or an event it must have belonged to those two classes. This algorithm was used to deal with the problems of binary classification, a method of class separation was used to distribute the classes differently and it has low complex issues in computation [41]. Quadratic discriminant analysis (QDA) method works for nominal labels and numerical attributes of objects. Discriminant analysis is used to determine the differences between two or more naturally occurring groups, it may have a descriptive or a predictive objective. Nearest neighbor method is used to select the metric measure which takes all the data from training for classification. Though it is a simple method yet has more complex issues and uses other algorithms to increase performance. K-Nearest Neighbor provides promising results than Nearest neighbor and its performance is also better for probability density functions, the value of K must be set by methods of validation as in [42]. Also, its computation complexity is high. Support Vector Machines (SVM) is based on the statistical learning method with a structured risk minimization principle [43]. SVM is used to classify the given input data from two predefined classes. The classifier predicts the object which largely depends on available data. SVM provides promising results in numerous cases of pattern recognition in [44]. It is not so accurate in most of the cases. However, a multiclass classification is used to fix the problems as in [45]. In vehicle classification only two classes exist, either it is a vehicle or not. In order to process this, we need a lot of data sets with different images of vehicles with different shapes. Thus the result is more complex and it was not always accurate in [45]. Margin maximization theory is used to rectify this problem by matching different views of vehicles in the dataset classes. SVM is praised as more robust tracking and detection algorithm, it is widely used in the aerial motion detection as in [46], [47]. In colored aerial videos, the objects are detected using Dynamic Bayesian Networks with the help of SVM. Neural Network classifier is widely used technique. The Multi-Layer Perceptron (MLP) also called an artificial neural network is one of the important classifiers. This technique uses the feed from input data which correlates with output data to form a graph. MLP consists of various layers of objects. Each node of a layer is connected to nodes of other layers in [48]. Adaboost Classifiers is used to develop or increase the performance of the classification techniques. If a classification technique is used to classify an object in terms of weights, an object with the lowest weight considered as unclassified and not properly formed, then its accuracy is also low. The technique creates an algorithm which increases weights of low accuracy and decreases the weight of high accuracy objects. This process is continued until a fine classifier is obtained. Adaboost uses various algorithms to enhance performance. Adaboost is weak with the data's outside the margin area but gives optimal results with the classifiers and has improved performance in [49]. SURF is mainly used for detection of features like recognition of objects, reconstruction of 3D, object registration and classification. It is derived partially from the descriptor of (Scale Invariant Feature Transform) SIFT. SURF is more effective, robust, and has high performance than SIFT. SURF is highly effective in tracking, detection, recognizing of objects and also it forms 3D scenes from those objects in [50]. HOGThe Histogram of Oriented Gradients (HOG) is a descriptive feature. In areas like Computer vision, and image processing uses this descriptor to detect objects. It extracts the gradient features from the object which is used for object detection in [51], [52]. ELVP Enhanced Local Vector Pattern (ELVP) is a vector based technique which expresses the structured data in given texture and 1D special structure with the corresponding pixels of the object. LVP is represented with the 2D directions of the available patterns with a diverse pattern of objects with respect to the reference pixel. Genetic Algorithm (GA) is used to enhance the methods of ELVP, SURF, and HOG. It creates genotype from the genome by populating strings and phenotypes are created by encoding the solutions of candidates. The results are of binary values such as 0s and 1s. GA begins to populate the results in random order by inversion and crossover operations in [52], [53].

| [8] | Lucas and Kanade | Image Registration Technique | The gradient feature of special intensity is used to match the objects in the images with the iteration of Newton-Raphson. | It can find the match even with fewer details and also can detect things even the object is rotated, scaled or sheared. | Image quality should be higher for matching of objects. |
| [25] | JM McHugh, J Konrad, V Saligrama, P Jodoin. | Adaptive background subtraction with Markov Random Field | Object detection is done by detecting the change in series of images. By MRF Spatial coherence is improved by change in labels of thresholds. | Achieving better performance by adapting statistical model, non parametric background model and MRF model to vary the thresholds. | When adding foreground with the background the detected regions are getting bigger rather than shrink. |
| [32] | Hsu-Yung Cheng, Chih-Chia Weng, Yi-Ying Chen | Enhanced Dynamic Bayesian Network | Automatic vehicle detection is proposed in this paper by using spatial and second derivative detection and tracking model | Objects are detected based on colors, shape and feature extraction intensity of pixels, then edge detection is done by Canny edge detector. | It cannot track multiple objects in videos of low quality |
### Table 2: Literature Survey on Vehicle classification

<table>
<thead>
<tr>
<th>Authors</th>
<th>Approach</th>
<th>Description</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>[54] Denis Kleyko, Roland Hostettler, Wolfgang Birk, and Evgeny Osipov</td>
<td>Vehicle Classification techniques Comparison by Machine learning on roadside sensors</td>
<td>The dataset of 3074 samples is processed for vehicle classification by D. Kleyko et al using different algorithms of machine learning. Various classification techniques are used such as SVM, neural networks and logical regression.</td>
<td>Logical regression shows the results had high performance when comparing with other methods of machine learning with the classification rate is 93.4%</td>
<td>The main difficulty in this method is the usage of datasets, as it was focused mainly on single class which is very difficult to search while classification.</td>
</tr>
<tr>
<td>[55] Zezhi Chen, Tim Ellis, Sergio A Velasit MIEEE</td>
<td>Comparison of vehicle type: Various Schemes of Classification</td>
<td>Vehicles are classified into four different classes car, bus, van and motorcycle. Two types of methods used here, SVM and random forest which is a feature.</td>
<td>The accuracy of SVM is 96.26% more robust than RF</td>
<td>Due to similar image size and shape of car, bus and van, miscalculation occurs.</td>
</tr>
<tr>
<td>[56] Muhammad Asif Manzoor, Yasser Morgan</td>
<td>SIFT Features is used to classify vehicle Make and its Model</td>
<td>The proposed method uses Linear SVM to process the data. The features are extracted using (SIFT )Scale Invariant Transform Feature derived by M. Mazoor et al [12]</td>
<td>The final result accuracy is about 89% against the NTOU-MMR dataset.</td>
<td>The front faced vehicle images are difficult to classify.</td>
</tr>
<tr>
<td>[57] A.H.S. Lai, G.S.K. Fung, N.H.C. Yung</td>
<td>Classification of vehicle type by visual based dimension estimation</td>
<td>The three classes of datasets used by A.H.S. Lai et al [13] are taxi, a mini-bus, and a double-decker. Estimation of vehicle is based on its length, width, and height</td>
<td>The accuracy obtained for estimation of vehicles is 92.5%</td>
<td>The main disadvantages faced are vehicles bumper close to road and vehicle mask.</td>
</tr>
<tr>
<td>[58] Seda Kul, Süleyman Eken, Ahmet Sayar</td>
<td>vehicle detection and classification in real time video streams</td>
<td>Distributed method of real time vehicle detection and classification system is proposed by Kul et al. [14]. Other techniques used here are vehicle classification, feature extraction, detection of foreground and background subtraction.</td>
<td>In broad daylight the results are promising with an accuracy of 89.4%</td>
<td>In night and bad conditions of weather they didn’t perform any work.</td>
</tr>
<tr>
<td>[59] Z. Dong, Y. Wu, M. Pei, and Y. Jia</td>
<td>Semisupervised Convolutional Neural Network is used for vehicle classification</td>
<td>In Dong et al. [15] method, Semisupervised Convolutional Neural Network is used while the classification of vehicles. The dataset consists of 9850 high resolution images are used. The dataset holds only front views of vehicles.</td>
<td>In daylight 96.1% accuracy is registered and in Night 89.4%</td>
<td>Misclassification occurs due to incorrect labels in the BIT dataset.</td>
</tr>
</tbody>
</table>
5 VEHICLE TRACKING APPROACHES

The object tracking in video processing system is a significant step to track the motion of objects in visual-based surveillance systems and it has been a challenging task for many researchers nowadays [60].

Figure 3: The Object tracking system.

The motion of objects such as vehicles has been tracked and located in dynamic scenes by the physical appearance of objects, to determine the movement of the blobs between two successive frames in the video analysis in [61]. Various vehicle tracking methods explained and concluded by many researchers for different problems are explained below:

5.1 REGION BASED TRACKING METHODS

In region based tracking methods, the particular region of the moving objects like vehicles (blobs) are tracked for locating the vehicles. These regions are segmented by subtracting the current image and previous image. A model-based automobile recognizing, vehicle tracking and classification are developed in [62] which is efficient and more reliable in several conditions. This method considered various positions and speed of moving vehicles until it is visible and worked on successive traffic scenes recorded by a static camera for automobile recognition. This region based model has three levels: raw data (original) images, region level, and vehicle level. The numbering of vehicles and vehicle classification in a traffic road management system has been illustrated in [63]. The flow of moving vehicles and classification of vehicles like car, bus, van are described and the scheme that removes false regions and shades elimination algorithm is achieved for more accurate and reliable segmentation of moving vehicles.

5.2 CONTOUR TRACKING METHODS

These methods depend on contours (the boundaries of vehicle) of vehicle in the process of vehicle tracking [64]. The video traffic surveillance system for real time supervision approach have proposed by the authors in [65]which makes use of optical flow and track a vehicle in 3D structure. This approach consists of two techniques: color contour based matching and gradient based matching, it produced more accurate results for tracking the moving vehicles, classification of objects, foreground and background detection of an object, vehicle flow, vehicle count, vehicle velocity and vehicle recognition. In a real time video surveillance of traffic, tracking and classification of vehicles have illustrated as in [66]. In this work, counting and classification of vehicles, detection of traffic lane change, direction and vehicle speed are detected. Multiple moving vehicles in the heavy traffic are detected and tracked even for various weather conditions, occluded objects like trees or shadows. For tracking and locating moving objects, Kalman filter, background subtraction methods, morphological processing operations are used for extracting and identifying the vehicle’s contour.

5.3 3D MODEL BASED TRACKING METHODS

These model based tracking methods involves occlusion detection, roadside trees, shadows in the moving vehicles and used a well known 3D model of the solid cuboid that suits to various types and sizes of vehicle images by varying the vertices to suit well as in [67]. By changing the region proportion, prototype width and height with respect to previous images, it achieves an efficient detection and tracking of vehicles. The classification of multiple moving vehicles such as low size vehicles (eg: bike, motorcycle), mid-size vehicles (eg: car, van, etc.), heavy vehicles (eg: trucks) are carried out in [68]. In this paper, the distance was measured by the use of 3D geometric shape of vehicles. A new framework of a 3D model-based tracking methods for vehicle detection and tracking based on a region proportion of boundary feature grouping was presented in [69]. This method has an advantage of more flexibility in detecting and tracking vehicles and achieves more reliable for many applications.

5.4 FEATURE BASED TRACKING METHODS

The feature based tracking method uses a feature descriptor of SURF (Speeded Up Robust Features) for a large region of feature sets to classify the vehicles in smart surveillance videos and it performed well in classifying similar and dissimilarity classes in [70]. A line-based shade method uses a linearity feature technique to remove all the shades in the occluded image. It also represented an automatic vehicle tracking and classification traffic observation system in [71].

5.5 COLOR AND PATTERN BASED TRACKING METHODS

The color and pattern of vehicle image series of traffic video surveillance is a technique used in [72]. This technique was used for the segmentation of foreground and background, vehicle flow, shade removal, vehicle velocity, vehicle count, vehicle location and this system is proved to work in different climatic conditions and is insensitive to lighting conditions. A real time traffic lane management is done by detecting and locating the vehicle to avoid congestion on roads, highways for the safety of the public [73]. There are three important main levels in this system that used are 1D shape patterns, tracking level, 2D pattern verification. Vehicle tracking plays a crucial role in the application of Intelligent Transportation Systems for the purpose of safety and security. Several algorithms used for vehicle tracking are mean shift algorithm, Camshift algorithm, optical flow, SURF, etc. Mean shift algorithm is a pattern matching algorithm which uses kernel function of histogram based targeted vehicle. It is used to track the vehicle efficiently when the blocked area of the targeted object is large [74],[75].
Cam Shift (Continuously Adaptive Mean Shift) algorithm is based on 1D histogram based object mainly for detecting faces and produces poor performance where the foreground object is same as background object or it varies color significantly in [74],[76]. SURF (Speeded Up Robust Features) algorithm is related to two dimensional Harr wavelet response and gives better solution than Camshift when the foreground object is same as a background object. As the computation time is high, this may not be suitable for tracking real time objects. Optical flow technique is used to differentiate multiple objects (foreground) and the background in an image. It is dependent on the distance between the movement of objects and a scene[77].

Kalman filter is used to track the multiple moving objects very effectively. The author proposed a framework to track the objects for a mobile robot traveling in crowded scenarios from deep tracking framework [78], [79]. In [80] the author developed outdoor tracking of moving vehicle based on deep learning framework. The features of the image are learned after pre-training a stacked denoising auto encoder and the next step is to add the k-sparse constraints to the stacked denoising auto encoder (Kssdae) then it is linked with a classification layer to enhance classification neural network. It is applied to an online tracker, after the process of fine tuning the evaluation produces a good performance of vehicle tracking after verification.

Table 3: Literature Survey on Vehicle Tracking

<table>
<thead>
<tr>
<th>Authors</th>
<th>Approach</th>
<th>Description</th>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>[62]</td>
<td>S. Gupte, et al.</td>
<td>Region-Based Tracking</td>
<td>The proposed system can detect, track, classify the vehicles with fewer data. It can also provide information about the detected objects such as its location and velocity.</td>
<td>This system insensitive to bad weather conditions, noise and variable illumination</td>
</tr>
<tr>
<td>[64] [65]</td>
<td>D. Koller</td>
<td>Contour Tracking</td>
<td>The proposed technique is fast and the results also quite promising.</td>
<td>The system is vulnerable to noise</td>
</tr>
<tr>
<td>Nelson H. C. Yung and Andrew H. S. Lai</td>
<td>3D Model-Based Tracking</td>
<td>In [47] the proposed model based tracking methods has described an occlusion detection process in the moving vehicles and used a well known 3D model of the solid cuboid that suits to various types and sizes of vehicle images by varying the vertices to suit well.</td>
<td>The system is very effective in the detection of vehicle occlusion.</td>
<td>The proposed system is not tested on vehicles in Traffic.</td>
</tr>
<tr>
<td>[70]</td>
<td>Xiaoxu Ma</td>
<td>Feature-Based Tracking</td>
<td>The proposed approach provides better performance.</td>
<td>When the view is changed the system is ineffective and occlusion also not tested.</td>
</tr>
<tr>
<td>W. Eric L. Grimson</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[72]</td>
<td>Seda Kul, Süleyman Eken, Ahmet Sayar</td>
<td>Color and Pattern-Based Tracking</td>
<td>The color and pattern of vehicle image series of traffic video surveillance are used for tracking. It consists of segmentation of foreground and background, vehicle flow, shade removal, vehicle velocity, vehicle count, vehicle location to track objects.</td>
<td>This system is proved to work in different climatic conditions and is insensitive to lighting conditions. The system needs to be tested under extreme weather conditions and occlusion problems also need to be checked.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6 APPLICATIONS
1. Monitoring Traffic conditions, pedestrian crossings, and parking areas.
2. Video surveillance for fire and smoke detection.
4. Monitoring vehicle speed.
5. Observation of day to day activities in shopping centers and amusement parks.

7 CONCLUSION
This paper provides a detailed study of the various techniques that are used in traffic video surveillance. It focuses on various techniques of vehicle detection, classification and tracking to make an efficient traffic management system by the use of video surveillance. Smart visual surveillance in dynamic scenes of various environmental conditions has been considered where outdoor environment is more challenging for researchers than indoor environment because of sunlight and illumination changes, human behavior, pedestrians crossing, anomaly detection of vehicles, waving of trees, shadows, lightning, etc. The overall study gives a better understanding and highlights the issues and solutions for traffic management systems.

REFERENCES
[55] Z. Chen, T. Ellis and S. A. Velastin "Vehicle type


[74] Sheldon Xu and Anthony Chang, Robust Object Tracking Using Kalman Filter with Dynamic Covariance, Cornell University.


