A Highly Robust Vehicle Detection, Tracking and Speed Measurement Model for Intelligent Transport Systems

S. Sri Harsha
Assistant Professor, Department of IT, VR Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India.

K. R. Anne
Professor, Director Academics, Veltech University, Chennai, Tamil Nadu, India.

Abstract
The high pace rise in vehicle counts, traffic density, and security concerns, a potential system for traffic surveillance, vehicle monitoring and control has become an inevitable need to facilitate intelligent transportation system (ITS). Although, numerous approaches have been proposed for moving vehicle detection and tracking, still the optimization needs can’t be ignored. In this paper, a robust image processing based vehicle detection, tracking and speed measurement model has been developed. The proposed model implements enhanced pre-processing, background subtraction, morphological operation as well as feature mapping processes for moving vehicle detection, tracking and speed estimation. To achieve optimal performance, a multi-directional filtering scheme has been develop for moving vehicle detection, which considers intensity, moving pixel orientation etc. for efficient candidate vehicle detection in traffic video. To enhance background subtraction, a novel multi-directional intensity strokes estimation approach has been introduced that plays significant role for distinguishing vehicle region from other background contents. In addition, the enhanced thinning and dilation based morphological process has been introduced that exhibits more precise and accurate vehicle detection. A novel feature clustering scheme with heuristic filtering based blob analysis and adaptive bounding box generation makes our proposed model more efficient for vehicle detection and tracking. Furthermore, a novel moving vehicle speed estimation approach has been developed that can be significant for efficient ITS systems.

Keywords: Vehicle detection and tracking, background subtraction, vehicle speed, intelligent transport system, directional filtering.

Introduction
The high pace emergence of communication system and associated technologies such as image processing has given a new dimension for Intelligent transportation systems (ITS), which has attracted attention of major researchers from industries as well as academics. The economic and easily availability of hardware have motivated researchers to develop more efficient solution for computer vision based applications. Image processing based computer vision has become a promising technology for real time supervision, monitoring, and control, that serves major areas, ranging from civil applications, industries till defence utilities. Especially considering significance of ITSs, the vision based supervision has the significant contribution, as it can facilitate real time monitoring, vehicle tracking and identification. In addition, other parametric identification such as vehicle speed, vehicle density, vehicle classification etc. has been demanded for efficient traffic monitoring and control. The provisioning for these requirements can enable optimal ITS solution in present day scenarios.

The exponential rise in vehicles on road has alarmed researchers to develop certain robust and efficient traffic monitoring system that could provide significant information about varied traffic related parameters such as vehicle count, traffic congestion and vehicle density, speed of vehicles etc. The fact or matter is that road accident caused deaths has increased with alarming rate globally. Interestingly majority of such accidents takes place due to uncontrolled the speed of the vehicle. Thus, an efficient vision based supervision framework for vehicle tracking and its speed measurement can be of great significance. A number of approaches have been proposed on the basis of video sequences processing and analysis. Some researchers have made efforts towards developing solution for moving vehicle/object tracking, traffic surveillance, traffic density estimation, trajectory estimation and vehicle speed measurement and most of the researches have used video processing and vision based concept [1-5] to achieve these objectives. However, majority of researches have either focused on vehicle or object tracking or its classification. On contrary, the irony is that there is an inevitable need for such systems which could provide optimal vehicle segmentation in all circumstances (environmental conditions, illuminations, background scenarios, noise etc.), without compromising with the accuracy of detection and classification. To ensure it, the algorithms require optimization at every level of video processing. In this paper, we intend to introduce novelty and optimization at each functional phase of image processing, segmentation, detection and tracking. There are numerous constraints such as conventional segmentation and background extraction, generic morphology approaches, light and brightness dependencies etc. In addition,
some other factors such as brightness, complexities in background, and speed of the moving vehicle etc are the predominant factors that cause intricacies in efficient detection and tracking. Specially, the consistent image brightness assumption between different sequential frames is one of the mostly used constraints. In order to optimize the vehicle detection accuracy and stability, some other constraints such as motion smoothness constraint has been suggested. However, the existing approaches could not solve the issues caused due to noise in ambience, disturbances, and occlusion. In real time scenarios, the background might spatiotemporally vary due to because of the natural complexity like non-uniform background texture, variation in background illumination, wind-driven motion etc. Furthermore, vehicle itself can exhibit a number of characteristics such as rigidity, elasticity or fluidity, which have significant impact of accuracy of detection and prediction. Hence, considering these existing intricacies and drawbacks of the existing approaches, in this paper, a video processing and vision based highly robust vehicle detection, tracking and speed estimation system has been developed. In order to enhance the efficiency of the proposed system, a number of optimization approaches have been implemented for efficient moving object or candidate (vehicle) detection, enhanced morphological approaches for precise detection, efficient blob analysis for noise reduction etc. that intends to enhance each components of the moving vehicle detection and speed estimation scheme. The other content of this paper is divided into five sections. In Section 2, the related works are presented, which has been followed by a brief of our contribution in Section 3. In Section 4, the proposed methodologies and its efficient implementation has been discussed. Section 5 discusses the conclusion of the proposed system. The references used in this paper are given at the last of the presented manuscript.

Related Works
In this section, some of the predominant literatures discussing vehicle detection techniques etc are discussed.

Motion Vehicle Detection and Segmentation
The detection of moving objects in the same image sequence, captured at different intervals is considered to be one of the interesting research fields in computer vision. There are a number of applications which consider detection of change as the event to perform various tasks such as surveillance system, computer aided diagnosis (CAD) and treatment, sensing system and monitoring and control for civil, industrial as well as military applications [6]. Traffic video surveillance deals with the analysis of the traffic images which contain moving vehicle detection and its accurate segmentation. A number of approaches have been proposed based on background substrations, frame differentiation and motion based algorithms for video processing [7, 8] but still these approaches can be found confined in dealing with vehicle detection and tracking, especially in case of dynamic scenarios. In general, this approach comprises three predominant mechanisms for vehicle segmentation. These are:
1. Background Subtraction approach.
2. Feature Based approach.
3. Frame Differentiation and Motion Based approaches.

In this paper, a hybrid model using background subtraction and feature extraction has been developed for vehicle detection, tracking and speed estimation. A brief discussion of the background subtraction scheme is given in following section.

Background Subtraction Methods
Background subtraction technique deals with extracting the moving foreground objects from background image. This approach has established itself as one of the extensively employed methods for change detection and can be used for vehicle regions detection. Still this approach suffers due to variations in lighting conditions and the climate situations [9]. Therefore, a number of researches have been performed to resolve these existing limitations of background subtraction methods. Researchers in [10, 11] proposed a statistical and parametric paradigm for background subtraction where they used the Gaussian probability distribution model for individual pixel in the image. In this approach, researchers have updated the pixel values using Gaussian probability distribution model iteratively from new image in a series to detect moving object. It has been followed by the categorization of the individual pixel (x, y) in the image where the pixels are categorized either be a part of the foreground or background. In [12], an enhanced approach was proposed for background subtraction where it was intended to segment moving vehicle from intricate and complex road conditions. They used a filtering approach on the basis of a histogram that gathers significant information from frame sequences with scatter background. Such background subtraction approach exhibited better in various conditions such as orientation, illumination, overcrowding situations etc.[12]. Researchers in [13] developed an example-based scheme for vehicle traffic detection. In the first step, they implemented an adaptive background approximation approach which was followed by division of frame into multiple non-overlapped sections to identify the candidate vehicles. In second step, researchers applied Principal Component Analysis (PCA) to estimate the two histograms for individual candidate, which was further used by support vector machine (SVM) for vehicle classification. Ultimately, the classified parts of the vehicle region were shaped and connected as a parallelogram so as to obtain the vehicle final shape. In fact, this approach is computationally highly complicate. Later, in [14] shadowing based vehicle detection was proposed, where the size features of vehicles were extracted. Researchers estimated the size as the distance between ends of front and rear tires for beneath vehicle’s shadow so as to differentiate the existence of vehicles on the lanes. A similar filtering based approach was proposed in [15], where researchers used two filters for eliminating swinging trees and raindrops. The proposed swinging trees algorithm has been used to reduce the computational complexity in vehicle tracking. Additionally, shade removal algorithm has been combined with a robust background deletion scheme to subtract moving vehicles from the background images. In [16], researchers developed a vehicle tracking algorithm by considering key features such as location, color dissemination, volume, speed of the group of the forefront entity and background based on Gaussian Mixture Model. Researchers used the concept to categorise individual pixel as noisy or forefront (background) so as to detect vehicle.
**Feature Based Methods**

In numerous researches, various structural features of vehicle such as, edges and corners have been used to segment it from background image [17, 18]. On the basis of these features they identified moving and static background for vehicle detection. In addition, researchers in [17] proposed a model to deal with occlusion problem between the overlapping vehicles, where they used the concept of background subtraction [17]. A background discrimination approach was developed using various image features and a trainable object detection approach was proposed in [18]. Initially, they used HAAR wavelet transform for feature extraction, which was then followed by learning based classification. They have used learning scheme that takes a set of labelled training data so as to perform labelling of the extracted objects features, as input. To perform classification, they used SVM classifier.

In [19], a sub region detection and analysis approach was used for feature extraction of the partially occluded vehicles. In addition, the used a multiscale transformation that posses the frame elements which are indexed by features such as orientation, position, measure and significant time-frequency localization characteristics, which have been retrieved through curvelet transform. The curvelet transform based feature extraction for vehicle detection algorithm was developed in [20]. In order to classify the vehicles, they used KNN and SVM classifiers. In [21], researchers developed a new traffic criterion detection paradigm where they used Epi-polar Plane Image (EPI). They used an enhanced Sobel operator based approach to deal with noise sensitivity and rough edge of the vehicles. In addition, to enhance performance, researchers used the Gabor operator texture edge detection scheme for feature extraction. In [22] a vehicle detection algorithm was developed, where they used the edges of the vehicle body, edge of windshield. The extracted features were processed for structural shaping using Bayesian network.

**Frame Differencing and Motion Based Methods**

This approach deals with distinguishing two subsequent frames so as to segment the moving object from the background image. In addition, this approach employs process to isolate moving blobs based on orientation and speed of the blob movement for detection optimization [23, 24]. Researchers suggested inter or intra-frame tracking levels to recognize and manipulate the occlusion vehicles. Researchers in [25] suggested for inter-frame and intra-frame manipulation for efficient vehicle detection and tracking [25]. A multimodal temporal panorama (MTP) approach was proposed in [26], where a remote multimodal (audio/video) monitoring system was developed to extract and reconstruct vehicles under moving conditions. Some other approaches have been proposed [27] which performs visual-based dimensional approximation for extracting motion vehicles. In addition, a shadow removal approach has been proposed to vehicle detection and classification [27]. In [24] a versatile movement histogram approach has been used to detect the moving vehicles. In their approach, a novel background variation model was used to enable lighting adaptation for vehicle detection in video data. In addition, researchers have used an adaptable movement histogram technique for vehicle detection.

**Our Contribution**

Detecting Moving vehicle from video accurately is challenging task. To detect moving object there are various approaches such as temporal differencing method, optical flow algorithm, background subtraction algorithm. Temporal differencing method uses two adjacent frames only to get background image. This method has one disadvantage that it cannot detect slow changes accurately. The approach of optical flow algorithm can detect object independently using camera motion, but unfortunately it is computationally complex and not suitable for real time application. In background subtraction absolute difference between background model and each instantaneous frame is taken to detect moving object. Background model is an image with no moving object. In our proposed model, a novel and enhanced background subtraction algorithm has been developed for moving vehicle detection. Our proposed model considers background subtraction concept for moving vehicle detection but unlike conventional approaches, we have introduced numerous algorithmic optimization approaches such as multi-directional filtering and fusion based background subtraction, thresholding, directional filtering and morphological operations for moving vehicle detection. The proposed system employs a directional filtering scheme for detecting moving vehicles, while considering its intensity and orientation variance as detection parameter. In addition, the multi-directional intensity strokes estimation approach has been applied that plays significant role for distinguishing vehicle region from other background contents. The implementation of the robust morphological scheme including thinning and dilation parameter with well calibrated content region identification makes our proposed system more robust and efficient. The feature clustering scheme with heuristic filtering based blob analysis makes our proposed model more efficient and precise for accurate moving vehicle detection. To enable better visualization of traffic monitoring, a bounding box generation scheme has been incorporated. In addition to the efficient vehicle detection system, in this paper a speed estimation mechanism has been developed, which measures the speed of vehicle in real time movement. The overall discussion of the proposed system is given in the following section.

![Functional Architecture of the proposed vehicle detection and surveillance system](image-url)
Proposed System
In this section of the presented manuscript, the proposed algorithms and its efficient implementation is discussed.

Video Data Acquisition
In this paper, in order to examine the performance of the proposed vehicle detection and speed estimation system for efficient traffic surveillance, the real time video data and some standard vehicle traffic data have been used. The real time video data has been recorded using camera with pixels adjustment facilities. The input video data are in RGB form, which are further converted into gray color format for processing.

Image Pre-Processing
To develop efficient vehicle detection and speed estimation scheme, the appropriateness of the input data and its quality is of great significance. In this phase, the pre-processing of the video has been done where the input RGB video has been converted into the frames that has been followed by the extraction of various parameters such as number of frames, frame rate, colour format, frame size etc. Unlike majority of existing systems, where the initial declaration of the total number of frames is must, in our proposed model, an automatic frame and dimensional extraction approach has been incorporated that makes our proposed system capable to process any kind of videos having different features and dimensional characteristics. Once retrieving the frames of the input videos, the RGB images (Figure 1) have been converted into gray color image (Figure 2), which is then followed by filtering and vehicle segmentation process.

Moving Vehicle Region Detection
Unlike conventional approaches, in this paper, it is intended to construct a feature map using multiple significant characteristic of the moving vehicle such as, vehicle edge strength, density, variance of orientations along with background subtraction scheme, as discussed in the previous section. Unlike majority of existing systems, where only the background extraction has been used as the foundation to detect vehicle, in our proposed vehicle detection approach a novel multilevel optimization approach has been introduced that ensures efficient video analysis and feature mapping for final video tracking purposes.

In our model, the resultant feature map is a gray-scale image having the same size of the input image, where the pixel intensity signifies the probability of Vehicle. The overall process of moving vehicle detection is discussed as follows:

Background Extraction
This is the matter of fact that the core of Background Subtraction approach is to retrieve the background of the moving video. In traffic surveillance system, while recording video on highway; it becomes highly intricate to get the image without any moving vehicle. In order to retrieve such image we have implemented background subtraction model. In this work, average of all frames pixel values, have taken into consideration. Thus, retrieving the background image, the region of interest extraction has been performed. In our proposed work, vehicle moving towards camera has been are tracked so that only one lane of road is considered as ROI. Each frame has been multiplied with extracted region of interest. In the proposed model, before processing for multiplication, the RGB frames have been converted into Gray level. Before processing for background subtraction, a motion integer background extraction has been done, where the background objects such as tree or other non-vehicle objects are eliminated to retain intact detection of moving vehicle (Figure 3).

In addition, a multi-directional filtering and fusion scheme has also been introduced in the proposed model, that along with above discussed background subtraction, assures optimal performance for precise background extraction and moving candidate region (vehicle) detection. The multi-directional filtering and fusion is discussed in next section. The proposed approach enables avoidance of any irrelevant movement such
as waving trees, or any other unwanted movement etc. It is required to do to get accuracy in vehicle detection. The absolute difference of each instantaneous frame and background model after multiplying both with extracted ROI has taken to detect only moving vehicles. The background subtracted frame is given in Figure 4.

![Background subtracted frame](image)

**Figure 4:** Background subtracted frame

### Threshold Estimation

Our proposed system employs a thresholding based segmentation scheme that converts grey scale image to binary image. This is the fact that the selection of an optimal threshold plays a vital role in assuring optimal image segmentation, especially in thresholding based segmentation. Therefore, in this paper, to distinguish foreground moving vehicle from the static background, thresholding scheme has been used. The considered conditional thresholding mechanism is given in the following equation (1).

\[
T(x,y) = \begin{cases} 
0 & \text{for } f(x,y) < S_{th} \\
1 & \text{for } f(x,y) \geq S_{th} 
\end{cases}
\]

where \(T(x,y)\) represents the threshold video frame, \(S_{th}\) depicts the selected threshold value and \(f(x,y)\) represents the instantaneous frame.

### Directional filtering

In order to achieve optimal performance, the magnitude of the second derivative of intensity has been used as a measurement of edge strength, because it facilitates optimal detection of intensity peaks that usually characterize vehicle in the current video frame. We have estimated the edge density of the moving vehicle on the basis of the average edge strength within a frame, which has already been converted from RGB video to the gray scale image. To enhance the vehicle detection efficiency, in this paper a novel multidirectional filtering has been proposed that estimates the variance of orientations in four directions 0°, 45°, 90° and 135°. Here, 0° indicates horizontal direction, 90° represents vertical direction, and 45° and 135° are the two diagonal directions. For simplistic implementation, we have used only horizontal and vertical directional filtering and fusion. The predominant advantage of this approach is that unlike conventional pixel by pixel and raw-column based scanning system, it performs image scanning in multiple direction simultaneously that as a result enhances computational efficiency and detection rate. In this paper, the convolution concept has been introduced with a compass operator (Sobel operator as depicted in Figure 5) that retrieves multi-directional edge intensity \((E_\theta=0or180,45,90,135)\) estimation of the moving vehicle frames. These all directional intensity vectors comprise all the characteristic of the edges of the frame including moving vehicle that enables effective vehicle detection and seed estimation.

![Compass operator image](image)

**Figure 5:** Compass operator

### Edge selection

This is the fact that vertical and horizontal edges can form the most significant strokes of the object (here moving vehicle) in an image and its lengths can also represent the dimensional characteristics of the corresponding vehicle, which can be significantly used to classify vehicles based on it geometry. Extracting and grouping these strokes, the vehicle region with different heights or dimensions can be located precisely. In practical scenarios, there can be both strong vertical as well as horizontal edges, reflecting the shape of vehicle. In addition, the horizontal and vertical edges generated by such moving objects can have large dimension, especially the length. Hence, performing classification of these edges into long and short edges can be significant to eliminate extreme large (vertical or horizontal) edges and the short edges can be considered for further processing for vehicle detection. Because of non-uniform background, color, intensity or illumination, long vertical edges generated by non-vehicle objects can have a large intensity and feature variance, such as pixel uniformity, color variations etc. Performing the thresholding process, such long vertical edges might turn out to be distorted short edges that as a result can cause false alarms. Similarly, non-uniform surfaces of the vehicle from various lighting, shadows and other features of the vehicle shape itself too cause broken vertical edges. To remove these false grouping caused due to those broken edges, in this paper, we have incorporated a two-stage edge generation approach. In the first approach the strong vertical edges are obtained as given in equation (1).

\[
E_{\text{vertical}}^{\text{strong}} = |E_{90}|_2
\]

where \(E_{90}\) represents the 90° intensity edge image which is nothing else but the 2D convolution result of the original image having 90° kernel, \(\cdot |_2\) represents a thresholding operator which is used to achieve a binary result of the vertical edges. As this approach intends to extract the strong edges, it can’t be stated to be sensitive to the threshold value. In the second approach, it is intended to retrieve the weak vertical as depicted through following equations (2-4):

\[
dilated = \text{Dilation}(E_{\text{verticalbw}}^{\text{strong}})_{3 \times 3}
\]

\[
closed = \text{Closing}(dilated)_{m \times 1}
\]
In this process, the morphological dilation has been introduced that plays significant role in eliminating the impacts of slightly slanted edges and a vertical linear structuring element \( m \times 1 \) which has been followed by the implementation of a closing operator so as to force the strong vertical edges clogged. There can be the trade-off on selecting the value of the size of structuring elements. Here, in this paper, we have assumed that the small value can be computationally efficient and consumes less time at the expense of false positives while a large value can significantly increase the precise detection but at the cost of elevated computational cost.

In our proposed model, considering the requirement of an effective and efficient system, \( m \) has been assigned as \( m = (1/25) \times \text{width}_{\text{frame}} \), which enables optimal vehicle detection results with an acceptable computation cost for a real time vehicle surveillance system. The ultimate edges formed are the combination of strong as well as weak edges, which has been retrieved using equation (5).

\[
\text{Edge}_{\text{Vertical} \_bw} = \text{Edge}_{\text{Vertical} \_bw}^{\text{strong}} + \text{Edge}_{\text{Vertical} \_bw}^{\text{weak}}
\]  

(5)

In our proposed model, a morphological thinning operator has been implemented which is followed by a connected component labelling algorithms. These functions are:

\[
\text{thinned} = \text{Thinning(Edge}_{\text{vertical} \_bw})
\]  

(6)

\[
\text{labeled} = \text{BWlabel(thinned,4)}
\]  

(7)

In our proposed vehicle detection and surveillance system, the morphological thinning operator makes the widths of the resultant edges one pixel thick and the connected component labelling operator performs labelling of the thinned vertical edges. In our proposed model, we have used 8 and 4-pixel connectivity for labelling. Performing labelling of the connected components, the individual edge has been uniquely labelled as a single connected component having its unique component number. Thus, the labelled edge frame has been processed by a length labelling process that intends to let the intensity of edge pixels represent their respective dimensions (lengths). Consequently, all the pixels belonging to the same edge have been labelled with the same number which is proportional to its dimensional length. Since, the high value in the length labelled vehicle video frame represents a long edge, in this paper a thresholding scheme has been employed to distinguish short edges (\( \text{short}_{\text{vertical} \_bw} \)). This is also the matter of fact that achieving 100% automatic precise vehicle detection in moving space in highly intricate task, in this paper, the efforts have been made to reduce the false negatives of missed detection. Here, we have used a low threshold value along with edge density and variance of orientation to optimize vehicle detection and precise speed estimation for efficient traffic surveillance system.

The outputs of the directional filters (vertical and horizontal) are given in Figure 6 and Figure 7. The combined vehicle detected is given in Figure 8.

**Feature Mapping**

In our proposed model, the practical facts that the regions with moving vehicle would have significantly higher edge density value, strength as well as variance of orientations as compared to the non-vehicle background regions. In our proposed model, these key characteristics have been exploited so as to enhance the vehicle region detection by means of generating a feature map values that significantly decreases the false regions and optimizes true candidate (moving vehicle) region detection. The overall process is illustrated through equation (8-10).

\[
\text{Candidate}_{\text{Vehicle}} = \text{Dilation(}\text{short}_{\text{Vertical} \_bw})_{\times m \times m}
\]  

(8)

\[
\text{Refined}_{\text{Vehicle}} = \text{Candidate}_{\text{Vehicle}} \times \sum_{\theta=90,180} E_{\theta}
\]  

(9)

\[
f_{\text{map}}(i,j) = N \sum_{m=-c}^{c} \sum_{n=-c}^{c} \{\text{Refined}_{\text{Vehicle}}(i+m,j+n) \times \text{weight}(i,j)\}
\]  

(10)

In this approach, the morphological dilation operator having a \( m \times m \) structuring element has been applied for selecting the short vertical edge image so as to get precise vehicle region detection. In our proposed vehicle detection system, multidimensional or multi-orientation edge information \( E_{\theta=90,180} \) have been used to refine (performing fusion of \( E_{\theta=90} \) and \( E_{\theta=180} \) ) the potential candidate moving vehicle detection. In equation (10), \( f_{\text{map}} \) represents the resulting feature map, and \( N \) gives a normalization operation that normalizes intensity (feature mapped values) in the range of \([0, 255]\). In our proposed model, a weight function \( \text{weight}(i,j) \) has been employed that estimates the weight of pixel \((i,j)\) on the basis of the number of orientations of edges within a video frame. Using the weight function, our proposed model distinguishes the candidate vehicle regions from background regions. The vertical \( (E_{\theta=90}) \) and horizontal scanning \( (E_{\theta=180}) \) outputs are given in Figure 6 and 7 respectively.

**Figure 6: Vertical Scanning**
Thus, on the basis of above vertical and horizontal scanning process, the vehicles have been detected using proposed motion vehicle detection scheme. Vehicles detected are illustrated through Figure 9 and Figure 10.

Now, combining vertical and horizontal detected vehicles, we get the combined vehicle output, given in Figure 10.

Feature clustering

The moving vehicles and its associated dimensional features can be clustered to localize the moving vehicle on the road. In fact, the characteristics of the components connected with moving vehicle are different from the static background. In practical scenarios, the characteristics such as the intensity of the feature map depicts the probability of vehicle in the current frame, certain simple thresholding can be applied to distinguish regions with higher vehicle possibility. Thus, employing certain morphological dilation operator the close regions can be connected together while ignoring or isolating those regions located far away. In our implemented vehicle detection model, a morphological dilation operator having square structuring element has been used that joints vehicle regions in the retrieved binary regions.
**Heuristic filtering Based Blob Analysis**

Being a highly sensitive application of traffic surveillance system, the filtration of retrieved region is of great significance. In this paper, we have applied a heuristic filtering scheme. The proposed heuristic filtering scheme possesses two constraints, which functions to filter out those blobs which do not contain vehicle regions. Here, the first constraint has been used to filter out all very small isolated blobs, where we have considered a relative area value rather than the absolute value for the individual blob. It enables our proposed system to function for the vehicle of any dimension or any size. In addition, our proposed model employs a second constraint that performs filtration of those specific blobs whose widths are very small as compared to the corresponding heights, because in realistic vehicles, usually height can’t be more than length or width of the vehicle. Thus, the proposed system can efficiently remove insignificant blobs to make prediction more accurate and precise.

**Boundary Boxes Generation**

Retaining the blobs reflecting vehicle in the running frame, it has been enclosed inside boundary boxes. We have estimated four pairs of the boundary box coordinates using the maximum and minimum coordinates of the top, bottom, left and right points of the subsequent blobs reflecting vehicle in the running frame. To avoid any probable missing of the vehicle related pixels existing near or outside of the initial boundary, width and height of the boundary box have been padded by a small amount. To make detection more precise, visible and road condition adaptive, in our proposed model, the large boxes such as borders, highway dividers etc. have been ignored and an additional adaptive padding has been introduced that makes our approach more accurate and efficient for better moving detection and tracking. The boundary boxes generated for each detected vehicle has been saved that makes tracking more efficient. The combined vehicle detected and its speed presentation is given in Figure 11.

**Vehicle Tracking**

In this paper, the proposed vehicle tracking system has been made on the basis of feature tracking concept. The features extracted have been tracked over sequential frames retrieved from input traffic video data. Unlike conventional approaches of tracking where researchers have used object matching algorithm based on Mahalanobis distance, in our proposed model, track identification and replica matching based tracking system has been developed. Here, initially the feature mapping for all frames has been estimated and a track graph has been prepared. To eliminate the probability of any error, few initial frames have been ignored. In this approach, a track has been deployed that traces the presence or passing of bounding boxes and thus indicating number of vehicles crossing the track. A search scheme has been used that searches bounding boxes in each frame and marks it for tracking status. The implemented function enables swift bounding box detection by means of a simultaneous horizontal and vertical search and match scheme. Detecting a bounding box while crossing the defined track, the vehicle has been counted and a template marking has been done that indicates the status of passing vehicle. Our proposed system represents an object matching scheme that estimates distance between vehicle features or the object features in the previous frame, which has been stored in track graph metrics and instantaneous frame. In addition, an additional marking template for vehicle ID presentation and speed estimation has been used that makes system better realizable and perceptible.

**Speed Estimation Scheme**

In this paper, the detected moving vehicle possessing its matching ID has been tracked over frames of the video data. In order to calculate the total number of frames having same object, has been estimated using following equation:

\[ \text{TotalFramesCovered} = \text{frame}_{\text{Last}, n} - \text{frame}_{\text{First}, 0} \]  \hspace{1cm} (11)

Where \( \text{frame}_{\text{First}, 0} \) represents the first video frame when vehicle enters into the region of interest (ROI) and \( \text{frame}_{\text{Last}, n} \) represents the frame when detected vehicle passed away from defined ROI. In addition, the real world distance has also been mapped on the image. Ultimately, the total frame count has been multiplied with the duration of one frame, which has been measured from the frame rate of video. In our proposed speed estimation model, with the fixed distance, the total time taken by vehicle to traverse it has been estimated in real world, which has been mapped into image.

The mathematical expression for speed measurement is given as follows:

\[ \text{Speed} = \frac{\text{Distance}}{(\text{TF} - \text{Framerate})} \]  \hspace{1cm} (12)

**Figure 11:** Vehicle tracking in speed measurement

As depicted in Figure 11, the speeds of the vehicles obtained are 15km/s, 16km/s and 11km/s.

**Conclusion**

The moving vehicle detection, tracking and its speed measurement system is of great significance for present day intelligent transport system. Considering limitations of the existing systems, such as conventional background subtraction, noise and illumination sensitivity, etc., in this paper a novel multi-directional filtering and fusion based background subtraction model has been developed that considers intensity,
Moving pixel orientation etc. for moving vehicle detection. The proposed multi-directional intensity strokes estimation scheme has enabled better performance for precise moving vehicle candidate detection and tracking so as to distinguish moving vehicle region from other background images. Further, the enhanced thinning and dilation based morphological process has made proposed system more robust and accurate. Performing moving vehicle detection, feature mapping has been done where feature clustering and heuristic filtering approach has been incorporated, which has made blob analysis more efficient to detect precise candidate vehicle region. Later, the boundary box generation has facilitated precise vehicle tracking. In addition, to the efficient moving vehicle detection and tracking, in this paper, an efficient vehicle speed estimation scheme has been developed that enables real time vehicle tracking and its speed measurement. Predominantly, this paper focussed on vehicle detection and tracking on single lane road that in future can be developed for multi-lane system. In addition, in future density estimation and vehicle classification can also be done.

Reference


A Robust and Efficient Optical Flow Analysis Based Vehicle Detection and Tracking System for Intelligent Transport System

S Sri Harsha, K. R. Anne
1Research Scholar, Acharya Nagarjuna University, Nagarjuna Nagar, Guntur,
2Assistant Professor, Department of IT, VR Siddhartha Engineering College, Vijayawada, Andhra Pradesh, India
3Professor, Director Academics, Veltech University, Chennai, Tamilnadu, India

ABSTRACT-In this paper, an enhanced optical flow analysis based moving vehicle detection and tracking system has been developed. A novel multidirectional brightness-intensity constraints (MBIGC) estimation and fusion based optical flow analysis (MDOFA) technique has been proposed that performs simultaneous pixel's intensity and velocity estimation in a moving frame for detecting and tracking the moving vehicle. The conventional Lucas Kanade and Horn Schunck optical flow analysis algorithms have been enhanced by incorporating a multidirectional BIGC estimation, which has been further enriched with a non-linear adaptive median filter based denoising. Such novelties have significantly enhanced the video segmentation and detection. A vector magnitude threshold based MDOFA algorithm has been developed for motion vector retrieval that eventually enables swift and precise moving vehicle segmentation from the background frame. A heuristic filtering based blog analysis has been applied for vehicle tracking. The MATLAB based simulation reveals that MDOFA-HS outperforms LK in terms of execution time and detection accuracy. In addition, the accurate traffic density estimation affirms robustness of the proposed system to be used in intelligent transport system.

Keywords: Multidirectional brightness-intensity constraint Optical flow analysis, intelligent transport system, Lucas Kanade, Horn Schunck.

I. INTRODUCTION

The exponential rise in technologies and associated applications has motivated researchers to develop certain efficient solution for a better living and security. On the other hand, low cost and efficient hardware availability has also introduced transition in computer vision based monitoring and control systems. Considering high pace rise in vehicle counts, traffic density and related concerns, a new scientific paradigm named intelligent transport system (ITS) has came into existence that intends to amplify the depth and width of vision based traffic surveillance system. The motion analysis and video processing based computer vision has emerged as a vital technique for real time traffic surveillance and timely reactive measures by security agents [1-5]. In addition, vision based vehicle detection, tracking, traffic density, vehicle speed, and classification can be of paramount significance for ITS decision process. The accidents caused due to high speed vehicle during overtaking have been alarming to strengthen ITS by enabling better vehicle detection, tracking and speed estimation systems. In last few years, numerous efforts have been made for video based vehicle detection and tracking. References [6, 7] examined various techniques for moving vehicle or object detection. However, their suggestions based on conventional background subtraction confines effective of the solution with varying traffic conditions, such as traffic density, number of vehicles, background features, vehicle color and geometry etc. Such feature complexity in real time traffic significantly influences the accuracy of moving vehicle detection and tracking. Background subtraction based approaches perform vehicle region segmentation based on the feature differences of the moving vehicle and background surfaces. On contrary, with the situations like night time, light condition or illumination, different weather conditions and non-linear road profile, spatiotemporal background variations (non-uniform background texture, illumination and wind speed) can significantly influence the detection accuracy. Majority of the existing approaches are not sufficient to deal with such situations because of high noise, disturbances, and occlusion. These approaches typically performs moving vehicles detection schemes operates based on the temporal difference between two consecutive frames [8,9], background subtraction [10] and optical flow estimation [11]. Researchers in [8] have applied optical flow analysis for vehicles navigation and object tracking [8].
The ability to perform moving pixel intensity detection makes optical flow analysis a potential alternative for moving object detection in complicate application scenarios. Optical flow analysis has emerged as a potential approach for video processing and motion analysis, especially for moving object detection. Optical flow analysis generates a 2D vector field that represents the motion field representing velocity and directions at each point of moving image sequence [12]. A number of constraints get introduced in between the frames to perform optical flow based motion analysis. These constraints function based on the image features such as pixel velocity and brightness. Predominantly, supposition of inter-frame image brightness dependability is one of the prime constraints. The implementation of optical flow analysis technique can be significant to alleviate the issues of occlusion and object overlapping that can enhance moving vehicle detection and tracking [13]. Conventional Lucas Kanade scheme was applied in [14] for moving object detection and suggested for further optimization. Researches [15, 16] dealt with stationary or dynamic object detection. Researcher [17] suggested for reduction in angular error caused in optical flow vectors of consecutive video frames. In this paper, an enhanced optical flow analysis based vehicle detection; tracking and density estimation system has been developed. Unlike conventional approaches, the proposed scheme introduces multi-directional brightness-intensity constraint (MDBIC) estimation and fusion based optical flow analysis (MDFOA) technique for vehicle detection and tracking. The performance of the MDBIC based Horn-Schunck (MDBIC-HS) algorithm has been applied for moving vehicle detection. To further enhance the performance, a non-linear adaptive median filter has been applied to denoise video input. It has significantly helped in highly accurate moving vehicle segmentation and detection accuracy. Additional, adaptive threshold based segmentation has been performed, which has been followed by heuristic filtering based blob analysis and vehicle tracking. In addition to the vehicle detection, the traffic density estimation has been performed.

The remaining sections of this paper are; Section II represents the MDFOA-HS and Lukas Kanade based optical flow analysis for vehicle detection and tracking. Section III discusses the results obtained, which has been followed by the discussion of conclusion and future scopes in Section IV. The references used in this research are given at the end of the manuscript.

II. OUR CONTRIBUTIONS

Considering the requirements of efficient vehicle detection and tracking system for ITS applications, various approaches have been proposed, in which background subtraction and optical flow analysis schemes are the predominant techniques. The proposed MDFOA based Horn-Schunck optical flow analysis technique has been developed for motion vector retrieval, which has been further used for moving vehicle detection. In addition, the conventional Lukas Kanade scheme has also been applied for vehicle detection and tracking. Unlike conventional Lukas Kanade and Horn-Schunck based optical flow analysis, we have used an adaptive median filter for speckle noise component’s elimination. Then, vector magnitudes thresholding based segmentation has been performed for detecting the moving vehicle in the video data. To further enhance the detection and tracking accuracy the blob analysis has been performed that remove the irrelevant pixels, thus enabling most accurate vehicle detection and tracking. The overall proposed System is illustrated in Figure 1.

A. Video Data Acquisition

In this research work, we have used the urban traffic surveillance video data to evaluate the performance of the proposed system. To perform vehicle detection the real time video traffic data has been obtained from a static camera placed at the road side. In our real time video retrieval, a static camera with auto pixel adjustment capability was mounted on the top of road. The input RGB video has been further processed into gray images to perform video processing and intended vehicle tracking.
Traffic Video Data Acquisition
Pre-processing (RGB to Gray)
Mean Estimation
Multi-Directional Fusion Based Optical Flow Analysis (MDFOA)
Adaptive Median Filtering Based Denoising
Vehicle Segmentation
Heuristic Filtering Based Blob Analysis
Vehicle Tracking
Vehicle Density Estimation

Figure 1: Multi-directional brightness-intensity constraint estimation and fusion based optical flow analysis (MDFOA) technique for vehicle detection, tracking and density estimation

B. Image Pre-Processing

The quality of video data and its appropriateness in terms of noise free input, and machine level data availability is of great significance. To meet these requirements, at first we have performed pre-processing of the input video data that enables input traffic data ready to process further. The initial processing has been performed by converting RGB to gray conversion and the initial process parameters such as the number of frames, frame rate, colour format, frame size etc have been obtained. Unlike majority of existing approaches where the prior dimensional declaration such as frame size and number of frames, etc is required, our proposed system performs automatic dimensional extraction that enables it to perform feature extraction and analysis with any input data. Due to the dynamic change in intensity and auto white balance feature of the camera, the mean of each video frame has been estimated on gray-scale format, which has been followed by optical flow analysis using our proposed MDFOA –HS and Lukas Kanade schemes for moving vehicle detection.

C. Multi-Directional Brightness-Intensity Constraint Estimation and Fusion Based Optical Flow Analysis

In general, optical flow analysis characterizes the trajectory and time rate of pixels in a time sequence of two consequential frames. Our proposed optical flow analysis technique operates with two dimensional velocity vectors (2D-V2) carrying significant information such as directional and velocity features in horizontal as well as vertical directions at certain point in an image (video frame). Since, in the proposed model, the directional filtered features such as brightness-intensity and velocity constraints are amalgamated together to characterize the pixels in the image, the proposed approach has been named as multi-directional fusion based optical flow analysis (MDFOA) scheme. The information retrieved from MDFOA are fused together to characterize certain point in terms of its intensity, velocity and brightness factors. In our proposed method, the 2D-V2 vectors have been applied on each pixel of the video data. The novelty of the proposed approach is its ability to perform information retrieval in horizontal and vertical directions simultaneously in the given image sequences. In our proposed model, the real time three dimensional (3D) input has been converted into equivalent two dimensional (2D) objects. Thus, estimating the 2D dynamic brightness functions, we have performed vehicle detection and tracking in the moving video. We the 2D functions have been applied to
perform brightness and velocity at certain location and distinct time instant \( I(x, y, t) \). The assumption that in the neighbourhood of an emigrant pixel, the variations in brightness and relative intensity doesn’t occur all along the motion field has been applied to estimate the brightness intensity function. Mathematically,

\[
I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t) \quad (1)
\]

Now, applying Taylor series on \( I(x + \delta x, y + \delta y, t + \delta t) \), the intensity vector has been obtained as

\[
I(x + \delta x, y + \delta y, t + \delta t) = I(x, y, t) + \frac{\partial I}{\partial x}(\delta x) + \frac{\partial I}{\partial y}(\delta y) + \frac{\partial I}{\partial t}(\delta t) + \text{Higher order term} \quad (2)
\]

Assuming Higher Order Term≈ 0 in (1) and (2), we get

\[
\frac{\partial I}{\partial x} \delta x + \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t} \delta t = 0 \quad (3)
\]

Now dividing (3) by \( \delta t \), we get

\[
\frac{\partial I}{\partial x} (\delta x/\delta t) + \frac{\partial I}{\partial y} (\delta y/\delta t) + \frac{\partial I}{\partial t} = 0 \quad (4)
\]

In other way,

\[
\frac{\partial I}{\partial x} (v_x) + \frac{\partial I}{\partial y} (v_y) + \frac{\partial I}{\partial t} = 0 \quad (5)
\]

Thus, the multidirectional brightness and intensity constraint at certain time instance \( t \) can be obtained as:

\[
I_x. v_x + I_y. v_y = -I_t \quad (6)
\]

In terms of the gradient constraints, the BIGC constraint has been derived as follows:

\[
\nabla I. \vec{v} = -I_t \quad (7)
\]

where \( \nabla I \) represents the spatial gradient of the brightness intensity factor and \( \vec{v} \) signifies the velocity vector of the optical flow of the image pixel. The variable \( I_t \) represents the time derivative of the brightness intensity gradient constraints (BIGC). In MDFOA based optical flow analysis techniques, the above derived equation (7) has been used to perform optical flow estimation. Equation (7) represents the BIGC that characterizes two unknown quantities using single function or equation. The concept that the gradient constraints function can significantly be used to perform optical flow based object detection and tracking, has been considered to perform vehicle detection and tracking. In this paper, we have applied the proposed MDFOA scheme with two different optical flow analysis methods; Lucas-Kanade (LK) and Horn-Schunck (HS), where these algorithms estimate the optical flow estimates. In our model, LK and HS algorithm only estimates brightness intensity function (1) and gradient constraints (7). A brief discussion of the proposed brightness intensity and gradient constraints (BIGC) schemes is given in the following sections.

a) Lucas-Kanade Model Based Vehicle Detection and Tracking

Win our proposed MDFOA and BIGC estimation, Lucas-Kanade (LK) method has been applied that introduces an error factor \( \rho_{LK} \) for individual pixel in the video-frame [12]. To estimate the error factor, the sum of the weighted least squares (WLS) of the gradient constraint (7) has been applied on each neighbouring pixels. Mathematically, \( \rho_{LK} \) has been obtained by:

\[
\rho_{LK} = \sum_{x,y \in \Omega} W^2(x,y) [\nabla I(x, y, t) \vec{v} + I_t(x, y, t)]^2 \quad (8)
\]
where \( \Omega \) represents the neighbouring pixels in the video frame; \( W(x, y) \) represents the weights of each neighbouring pixels (\( \Omega \)) in the moving frame. To ensure minimal error \( \rho_{lK_{\text{Min}}} \), the error factor \( \rho_{lK} \) has been estimated by each elements of the velocity vector while keeping result as zero. Finally, the optical flow output has been obtained in terms of a matrix given by

\[
\tilde{v} = [A^T W^2 A]^{-1} A^T W^2 \tilde{b}
\]  

(9)

Further, to estimate the values of \( A, W \) and \( b \) for \( N \) neighbouring pixels \((n \times n)\) neighbouring pixels (i.e. neighbour pixels of \( \Omega \), \( where N = n^2 \)) and \((x_i, y_i) \in \Omega \) at certain time instant. the following mathematical functions have been applied:

\[
A = [\nabla I(x_i, y_i), ..., \nabla I(x_N, y_N)]
\]  

(10)

\[
W = d\text{iag} [W(x_i, y_i), ..., W(x_N, y_N)]
\]  

(11)

\[
\tilde{b} = -[l_i(x_i, y_i), ..., l_i(x_N, y_N)]
\]  

(12)

Now, putting the respective values of \( A, W \) and \( b \) the velocity for each pixel in the frame has been obtained \( (12). \) Unlike conventional summing up based optical flow analysis methods \([13] \), in this paper Gaussian or the differential temporal gradient filter (DTGF) based convolution technique has been used. This approach significantly reduces the computational complexities in fusing the BIGC at different instants. Applying the (DTGF) based convolution technique the multi-directional fusion has been done. Unlike conventional Lukas Kanade (LK) based in this paper a dual error function based method called Horn-Schunck (HS) has been used. In order to evaluate the performance of the proposed MDFOA based optical flow analysis, in this paper the proposed MDBC has been applied with both LK algorithm as well as HS optical analysis scheme \([14] \). Unlike LK based scheme, a dual error function based HS method has been applied with the proposed MDFOA approach. A brief discussion of the proposed scheme is given as follows:

**b) Horn-Schunck Model Based Vehicle Detection and Tracking**

LK optical analysis method applies a single error function \( \rho_{lK} \) to estimate BIGCs, but considering high dynamics caused due to fast moving vehicles, it is found confined. In addition, it becomes time consuming. Therefore, to deal with such issues, in this paper, the proposed MDFOA based HS scheme has been applied that introduces an additional error factor called “global smoothing factor”. It enables our proposed system to deal with extreme dynamism and variations in the optical flow vector elements \((v_x, v_y)\) in the neighbouring pixels \( \Omega \). In order to reduce the total error \( \rho_{HS} \) of the proposed MDFOA-HS approach, the following mathematical equation has been applied:

\[
\rho_{MDFOA-HS} = \int_D (\nabla I \cdot \tilde{v} + I_t) + \lambda \left[ \left( \frac{\partial v_x}{\partial y} \right)^2 + \left( \frac{\partial v_y}{\partial x} \right)^2 + \left( \frac{\partial v_x}{\partial x} \right)^2 + \left( \frac{\partial v_y}{\partial y} \right)^2 \right] dxdy
\]  

(13)

where \( D \) represents the complete frame region or complete region in each image of the surveillance video, \( \lambda \) states for relative effect factor of the second introduced error term. Here, \( \lambda = 1.0 \) has been considered. Thus, deriving \( \rho_{MDFOA-HS} \), it becomes feasible to apply Jacobian model or the Gauss-Seidel iterative methods \([15] \) to model the system for performing optical flow analysis based vehicle detection and tracking. This is the fact, that conventional HS approach delivers higher accuracy even with the higher vehicle density conditions and extreme movement (here it is important because in highway traffic model, there can be very fast moving vehicles), but it requires more iterations to perform overall BIGC estimation, and hence it is relatively slower as compared to LK scheme. The HS method has been applied for BIGC estimation between current frame and \( n^{th} \) frame of frame sequences that enables robust functioning and accurate tracking of vehicle. Once performing BIGC estimation and optical flow analysis in each frame of the traffic surveillance videos, the noise filtering has been applied using non-linear adaptive median filter. A brief of the noise filtering process is given as follows.
D. Noise Filtering

In general, the moving digital images are influenced by a number of noise distributions that primarily depends on the functional conditions. These noise components can be impulsive, additive or certain signal dependent noise components and even the amalgamation of these all noises [13]. In numerous cases, the noise intensity gets exponentially increased due to change in image features such as intensity, contrast background etc in video surveillance images. Such noise components might even deteriorate the pixels and thus turning same pixel with varied intensity levels as compared to the neighbouring pixels [14]. MDFOA scheme estimates BIGC factor for each moving frame so as to detect and track moving vehicle. In such cases, the corruption or degradation in uniformity of the pixel intensity might lead inaccurate and degraded performance. To alleviate such limitations, in the proposed work a non-linear adaptive median filter has been applied that effectively denoises each frame of the traffic surveillance video for further processing. Thus, introducing this denoising technique, the suppression of noise components in homogenous regions has been performed. In addition, the proposed model performs spatial as well as temporal edge feature conservation along with the elimination of random impulses that as a result enables noise free frame for further video processing. Here, the proposed non-linear adaptive median filter has been applied on $3 \times 3$ adjacent neighbouring pixels where it substitutes the value of a pixel by the median of the gray levels of the neighbouring pixel in adjacency. To preserve the edge information and other significant information for video process, anisotropic diffusion (ASD) can also be applied.

E. Image Segmentation

In order to perform vehicle detection in traffic surveillance video, the video frames have been segmented into certain concept region using adaptive threshold estimation based background subtraction. The obtained optical flow vectors have been applied to determine whether the pixels in the current frame belong to the moving object or not. The adaptive threshold estimation has been performed over resulting optical flow vectors that significantly distinguishes the moving concept region or ROI (vehicle) from the background. Unlike conventional approaches, the proposed adaptive thresholding scheme varies from one frame to other. Different features like color, contrast, illumination, background intensity and camera calibration etc have been used as the spatio-temporal features to perform adaptive thresholding process. To perform adaptive thresholding, the following equation has been used.

$$T_h = \left( |u|^2 + |v|^2 \right)^{1/2}$$

(14)

In (14), $T_h$ represents the absolute value of the optical flow and the respective threshold is estimated using the absolute values of $T_h$. The segmented concept region or the ROI (moving vehicle) has been obtained using morphological closing operators of MATLAB image processing toolbox. Performing morphological closing function, the holes and relevant pixels are connected together so as to preserve the vehicle shape and appearance. We have applied the following morphological closing function on the structural element $S$.

$$T \cdot S = (T \oplus S) \oplus S$$

(15)

where,

$$S = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

(16)

Here, the matrix $T$ contains the information about the moving vehicle, which is retrieved by means of thresholding based segmentation. The segmented region is further applied to perform moving vehicle tracking in running video’s frame sequences. The proposed thresholding based vehicle tracking scheme meets the following thresholding criteria (15).

$$T_h = \begin{cases} 255, & \text{if } T_h > \text{Defined Threshold (0.039)} \\ 0, & \text{Otherwise} \end{cases}$$

(17)
During this process, the predominant intricacies observed was the segmentation of the pixel blocks having very minute size that creates ambiguity to be or not to be a part of ROI concept region. To alleviate these issues, the blob analysis has been performed that significantly suppresses the irrelevant blobs to ensure optimally segmentation.

**F. Heuristic Filtering Based Blob Analysis**

In order to eliminate the irrelevant and insignificant blobs from each video frame, we have applied an additional heuristic filtering that encompasses two constraints. These heuristic constraints remove the blobs containing no vehicle regions. The first constraint filters out all very small isolated segmented regions or blobs, where a defined region has been considered as reference for each blob. In this paper, 3 pixel connectivity based blob analysis has been performed, where for individual frame, the statistics of the neighbouring connected components (3x3 pixels) have been obtained. It is then followed by the generation of \( 4 \times N \) matrix representing the bounding box coordinates. Similarly, a matrix of \( 2 \times N \) has been generated to represent the centroid coordinates, where \( N \) states the number of blobs. Furthermore, image arithmetic functions such as image addition and subtraction have been applied to achieve a binary image with only centroid. Finally, the output video has been converted into frames, which has been further processed to retrieve the matrix with 2D centroid coordinates. In this research, to further simplify the blob analysis, we selected the blobs with the fixed dimension of \( 300 \times 3000 \) pixel. It has enabled our system efficient to perform detection with any geometric dimensions. Meanwhile, the second constraint performs filtering of those particular blobs having relatively very small width than corresponding heights. This is because, in real application scenarios, height can’t be more than length or width of the vehicle. In such manner, the vehicle concept region or the ROI has been identified for further tracking purposes.

**G. Boundary Boxes Generation And Tracking**

Performing the blobs analysis, the detected vehicle has been enclosed within a boundary box. Here, four pairs of the boundary box coordinates along with a centroid coordinate have been applied to represent the subsequent blobs representing vehicle in the running video. To make detection more precise, visible and road condition adaptive, the large boxes such as borders, highway dividers etc have been ignored and an additional adaptive padding has been introduced that makes our approach more effective for moving vehicle detection and tracking.

### III. RESULTS AND ANALYSIS

In this paper, the proposed MDFOA based optical flow analysis based vehicle detection and tracking system has been developed using MATLAB/SIMULINK software with image processing and Vision toolbox. To evaluate the performance of the proposed systems, different traffic surveillance videos data have been used. Initially the input raw videos have been converted from RGB to gray image, which has been further processed for BIGC estimation and MDFOA based optical flow estimation. Here, the proposed MDFOA scheme was applied with Horn Schunk based optical flow analysis algorithm. It was then followed by the processes such as non-linear adaptive median filter based filtering, adaptive threshold based segmentation and heuristic based blob analysis. To compare the performance the proposed MDFOA-HS and LK was also developed for moving vehicle detection. The developed models have been simulated on Windows-7 OS with Dual core processor, 4 GB RAM, and 1.8 GHz processor. The comparative performance of the proposed MDFOA-HS and LK based scheme is summarised in Table 1.

<table>
<thead>
<tr>
<th>Performance parameters</th>
<th>LK Based Vehicle detection (Second)</th>
<th>MDFOA-HS Based Vehicle detection (Second)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total recorded time</td>
<td>13.856</td>
<td>12.76</td>
</tr>
<tr>
<td>BIGC and optical flow estimation</td>
<td>2.343</td>
<td>3.968</td>
</tr>
<tr>
<td>Threshold estimation</td>
<td>0.400</td>
<td>0.109</td>
</tr>
<tr>
<td>Morphological closing</td>
<td>0.265</td>
<td>0.265</td>
</tr>
<tr>
<td>Blob analysis</td>
<td>0.156</td>
<td>0.078</td>
</tr>
<tr>
<td>Velocity</td>
<td>0.140</td>
<td>0.078</td>
</tr>
<tr>
<td>Median filter</td>
<td>0.812</td>
<td>0.843</td>
</tr>
</tbody>
</table>
Accuracy

|          | 93.81% (91/97 vehicles) | 98.96% (96/97 vehicles) |

Observing Table 1, it can be found that the proposed MDFOA-HS based vehicle detection and tracking system performs better than conventional Lukas Kanade (LK) optical flow analysis algorithm. These results have been obtained for a defined and equal simulation period. Figure 2 represents the original traffic surveillance video frame, which has been further processed with MDFOA-HS approach. The results obtained for motion vectors is presented in Figure 3. Figure 4 represents the adaptive threshold based segmentation. Here, the robustness of the proposed adaptive filtering and heuristic based blob analysis can be observed easily. The bounding box based vehicle detection and tracking can be found in Figure 5. To perform accuracy analysis, the number of vehicles detected by the proposed algorithms has been compared with the manual calculation. The results reveal that MDFOA-HS based vehicle detection outperforms conventional Lukas Kanade (LK) based optical flow analysis in terms of detection accuracy. Interestingly, it has been observed that the MDFOA-HS based scheme takes bit higher time in computation. It can be because of the computational overheads caused due to multidirectional filtering and fusion based BIGC estimation as well as due to double error factor estimation. In addition, the proposed system has been examined for its effectiveness in estimating the traffic density, where the proposed system has outperformed conventional LK based vehicle detection and tracking system. Thus, the overall results reveal that the proposed MDFOA-HS scheme can perform better for vehicle detection and tracking than its counterparts.

IV. CONCLUSION

Considering, the requirement of novel vehicle detection and tracking system for intelligent transport system (ITS), in this paper, a novel and robust multidirectional filtering and fusion based optical flow analysis (MDFOA) scheme has been developed, which has been implemented with Horn Shunck (HS) optical flow algorithm. The proposed scheme encompasses varied novelties in terms of enhanced brightness and intensity gradient constraints (BIGC) estimation, non-linear adaptive noise filtering, heuristic filtering based blob analysis adaptive threshold based segmentation and bounding box generation for vehicle tracking. The implementation of simultaneous velocity and intensity estimation
at each pixel enables the proposed system efficient. Retrieving the BIGC features, the motion and velocity vector components have been obtained which has been further applied to perform adaptive thresholding based segmentation. This novelty has enabled the proposed system to deliver optimal detection of moving vehicle. The heuristic filtering based blob analysis has exhibited efficient performance in reducing unwanted blobs from the video frame, and thus resulting into enhanced vehicle detection and tracking accuracy, even at high speed vehicle movement. Performing the boundary box generation, the tracking of the vehicle has been done. In addition, the vehicle density estimation too has been done based on their crossing frequency through a defined area in the frame. The comparative results between MDFOA-HS and Lukas Kanade based vehicle detection affirms better results by the proposed system. The detection accuracy of 98.96%, with relatively appreciable time efficiency affirm that the proposed MDFOA-HS based scheme can be used for high speed moving vehicle detection and hence can be a potential technique for ITS utilities. In future, the effectiveness of the proposed scheme can be examined for night time vehicle detection and tracking and even certain vehicle classification model can also be explored.

REFERENCE

Gaussian Mixture Model and Deep Neural Network based Vehicle Detection and Classification

S Sri Harsha
Assistant Professor, Department of IT,
VR Siddhartha Engineering College, Vijayawada
Andhra Pradesh, India

K. R. Anne
Professor, Director Academics,
Veltech University, Chennai,
Tamil Nadu, India

Abstract—The exponential rise in the demand of vision based traffic surveillance systems have motivated academia-industries to develop optimal vehicle detection and classification scheme. In this paper, an adaptive learning rate based Gaussian mixture model (GMM) algorithm has been developed for background subtraction of multilane traffic data. Here, vehicle rear image information and road dash-markings have been used for vehicle detection. Performing background subtraction, connected component analysis has been applied to retrieve vehicle region. A multilayered AlexNet deep neural network (DNN) has been applied to extract higher layer features. Furthermore, scale invariant feature transform (SIFT) based vehicle feature extraction has been performed. The extracted 4096-dimensional features have been processed for dimensional reduction using principle component analysis (PCA) and linear discriminant analysis (LDA). The features have been mapped for SVM-based classification. The classification results have exhibited that AlexNet-FC6 features with LDA give the accuracy of 97.80%, followed by AlexNet-FC6 with PCA (96.75%). AlexNet-FC7 feature with LDA and PCA algorithms has exhibited classification accuracy of 91.40% and 96.30%, respectively. On the contrary, SIFT features with LDA algorithm has exhibited 96.46% classification accuracy. The results revealed that enhanced GMM with AlexNet DNN at FC6 and FC7 can be significant for optimal vehicle detection and classification.

Keywords—Vehicle detection and classification; deep neural network; AlexNet; SIFT; Gaussian Mixture Model; LDA

I. INTRODUCTION

The high pace development of technologies predominantly image or video processing techniques have enabled a number of application scenarios. Visual traffic surveillance (VTS) based intelligent transport system (ITS) is one of the most sought and attractive application and research domains, which has attracted academia-industries to enable better efficiency. The significant application prospects of ITS systems have motivated researchers to achieve a certain effective solution. An efficient vehicle detection and localization scheme can enable ITS to make efficient surveillance, monitoring and control by incorporating semantic results, like “X-Vehicle crossed Y location in Z direction and overtaking A Vehicle with Speed B”. Considering these needs, in previous works [1,2], vehicle detection, tracking, and speed estimation model were developed. However, the further optimization could enable more efficient ITS solution. Developing a novel and robust system to detect and classify the vehicle simultaneously can be of paramount significance. Vehicle detection features such as size, shape, color, stopped or moving object and their type can be vital for ITS decision systems [3]. The type of the detected vehicle can provide significant information that may lead ITS administrators to ensure that certain type of vehicle doesn’t appear in a certain region. Implementation of the multi-camera infrastructures [4] might enable ITSs to identify and detect the targeted vehicle or vehicle class by matching it from traffic data from different functional data acquisition cameras.

Recently, some efforts have been made for vehicle detection and tracking, however, very few efforts have been made towards its classification. Especially, not much effort has been made on developing a simultaneous vehicle detection and classification system. There are numerous issues like cluttered image scene, occlusion, the exceptionally higher number of classes and features, etc. that make classification highly intricate. Background segmentation based object detection can be beneficial as it can remove clutter [5]. The images retrieved through surveillance cameras are used to be of low resolution, different lighting conditions, and more importantly, size of the vehicle is very small in complete traffic video frame that makes classification too tedious task. In practice, vision-based surveillance applications require dealing with huge unlabelled data elements, features, occlusion, unannotated images, localization and classification under different lighting or background conditions, etc. To deal with such issues, a system with effective background subtraction, feature extraction and vehicle region or ROI localization and classification can be of paramount significance. Also, to provide time effective solution reduced data processing and computationally efficient approach is required. Considering such requirements and motivations, in this paper, a multilevel optimization measure has been proposed. In this paper, an enhanced Gaussian Mixture Model (GMM) algorithm and connected component analysis (CCA) scheme has been developed for optimal vehicle region or ROI identification and localization. Further, to enable accurate and swift vehicle classification an enhanced multilayered deep convolutional neural network (DNN) was developed that functions on AlexNet DNN model. An additional feature extraction model, space invariant feature transform (SIFT) were prepared to extract ROI features. Implementing dimensional reduction schemes over extracted features support vector machine (SVM) based classification was performed that classifies vehicles into different classes.

The other sections are divided as follows: Section II presents the related work. Section III discusses the proposed research or contribution. In Section IV, the algorithmic development and its discussion are presented and the results
obtained are given in Section V. Section 6 presents conclusion and future work. References used are presented at the last of manuscripts.

II. RELATED WORK

Recently a vision based model for vehicle detection, feature extraction and classification were developed in [6], where researchers applied GMM with Hole Filling algorithm for vehicle detection. Gabor kernel based feature extraction and Multi-Class classification. To deal with dense vehicle classification, a vector sparse coding scheme with SVM was proposed in [7]. Applying sparse coding technique, they projected features to the high dimensional vector that assisted SVM to perform better classification. The combined shape and gradient feature based classification was proposed in [8][9]. To perform shape-based classification, at first they performed background subtractions and obtained shape features from silhouettes in the omnidirectional video frames. Similarly, for gradient based classified Histogram of Oriented Gradients (HOG) Features were obtained, where researchers found that the combined features based classification can be more useful than the individual feature based classification. The features like geometry, number plate location and shape was used as input of dynamic Bayesian network (DBN) for vehicle classification [10]. Researchers applied GMM to calculate the probability distribution of features. However, they could not address the detection issues under varying illuminations and frame dynamicity. A sparse learning based vehicle detection and classification model were proposed in [9]. Later, in [11] sparse coding and spatial pyramid matching scheme were used for vehicle classification, where they extracted the patch based sparse features using a discriminate dictionary. The extracted features were classified using histogram intersection kernel based SVM classifier. An integrated vehicle detection and classification model was proposed in [12] where multi-resolution vehicle recognition (MRVR) scheme was introduced to support cascade boosted classifiers for vehicle classification. The combined feature including HAAR and HOG was used for vehicle detection and classification [13]. The concept of multi-feature fusion was proposed in [14], where authors combined local as well as global feature of the detected vehicle region or ROI. In their work, they applied SIFT for local feature extraction and PCA based global feature extraction process. The combined features were used for classification using SVM [14]. To increase accuracy, researchers [15][16] used higher layer features of the deep neural network (DNN). Researchers [15] extracted PHOG and LBP-EOH using DNN. They combined these features for classification. An appearance based vehicle classification scheme has been developed in [17], where vehicle front features has been applied for classification using semi-supervised CNN algorithm. On the contrary, in this paper, the rear information and lane dash line information have been applied to perform multi-lane vehicle detection. Also, it deals with occlusion issues. A shape-based multi-class classification scheme has been proposed in [18] where the concavity property of vehicles such as buses and sedans was used for classification. Authors in [19] applied a Deep Belief Networks (DBN) based vehicle classification. They have used key features such as image pixel value, HOG features and Eigen features to perform classification. An approach named cascade classifier ensemble has been suggested in [19] for vehicle classification. As the first ensemble, they applied classifiers such as SVM, K-NN, random forest and multiple-layer perceptrons (MLPs) for vehicle classification.

Recently, real-time vision-based vehicle detection and the classification system were proposed in [20], where a simple morphology-based approach has been formulated for ROI detection. To deal with vehicle occlusion issues, they applied the ROI accumulative curve method and Fuzzy Constraints Satisfaction Propagation (FCSP). Retrieving the Time-Spatial Images (TSI) from the surveillance video, they eliminated shadowed region using SVM and Deterministic Non-Model Scheme (DNMS). A combined model to perform vehicle detection, tracking, classification, counting has been proposed in [21]. In [16], researchers applied conventional median filter and Otsu method based background subtraction for vehicle detection. However, they could not address the problems introduced due to illumination change and background features variations. To deal with these issues, GMM scheme can be a potential alternative for background subtraction [6][10], however, traditional GMM scheme remains questionable especially with dynamic frame movement and varying illumination conditions because of its fixed learning rate and pixel saturation issues. To deal with this in this paper, an adaptive learning rate based GMM model has been developed for vehicle ROI detection. On the other hand, the direct deep neural network (DNN) implementation for vehicle detection and classification is highly intricate and almost impractical. Therefore, in this paper an enhanced AlexNet DNN with CaffeNet model [22] has been developed that enables optimal vehicle detection and classification, even with huge dataset. Considering the effectiveness of the SVM classifier, in this paper, 10-fold cross-validation scheme has been applied to achieve accurate classification performance.

III. CONTRIBUTION

In this paper, a robust vehicle detection and classification system has been developed for vision-based surveillance system to be used for ITS purposes. In fact, the presented work is a multilevel optimization effort where numerous optimization efforts have been introduced on a different phase of vehicle detection and classifying. The proposed approach includes enhanced GMM (adaptive learning rate) based background subtraction and vehicle detection, CCA based ROI identification or localization, DNN model; AlexNet and CaffeNet based feature extraction, dimensional reduction and SVM based efficient vehicle classification. To perform vehicle localization in image and occlusion avoidance, the vehicle’s rear features along with lane dash markings have been applied. Once performing background subtraction, to reduce irrelevant blob presence, CCA has been applied that eventually achieves precise vehicle region or ROI. To extract ROI features, an enhanced DNN algorithm has been applied based on convolutional neural network (CNN) principle. Here, AlexNet DNN model [23] extracts multidimensional features at the higher DNN layers (Fig. 3). In existing works [23], DNN has been used for vehicle classification using different datasets [24]. However, AlexNet can't be applied directly as in practical situations the labeled data used to be smaller than the DNN parameters. In generic DNN based approaches the probability
of degraded accuracy and over-fitting can’t be ignored. To deal with this issue, in this paper, CaffeNet [22] with AlexNet DNN has been used that enables optimal performance even with general purpose computing systems. In practice, due to higher unannotated data, performing DNN learning and classification is a tedious task. To deal with such issues, multilayered DNN has been implemented and trained over large scale labeled vehicle dataset that enables swift and accurate data classification. In this work, the ROI features have been retrieved at each layer of the trained DNN (Convolutional Layer-1 to Layer-5 and Fully Connected Layer 6 and Layer 7). Since, features at the higher layers (fully connected 6, 7 and 8) of DNN used to be more informative [16] and therefore a set of 4096-dimensional features have been retrieved for individual vehicle image at FC6 and FC7 (Fig. 3). Recently, researchers [25] suggested that SIFT features can also enable accurate classification; therefore in this paper, 4096 SIFT feature descriptors have been obtained from each image, which is equivalent to AlexNet FC6 and FC7 features. The extracted features have been projected to the dimensional reduction schemes, the principle component analysis (PCA) and linear discriminant analysis (LDA). After dimensional reduction with PCA and LDA individually, the retrieved AlexNet features have been projected to the polynomial kernel based SVM classifier for vehicle classification. Similarly, SIFT feature vectors have been used as input of SVM for classification. The detailed discussion of the proposed vehicle detection and classification system is presented in the following sections.

IV. SYSTEM MODEL

This section discusses the overall development and implementation of the proposed enhanced GMM and DNN based vehicle detection and classification system (Fig. 1).
The overall proposed model comprises four sequential phases. These are:

A. Vehicle detection,

B. Feature extraction,

C. Dimensional reduction and

D. Classification

The discussion of the proposed methodology is presented as follows

A. Vehicle Detection

This section discusses the proposed vehicle detection mechanism.

1) Multilane road image retrieval: In this work, the vehicle image data has been obtained using static a camera placed on the road side. In real-time vision based surveillance applications, occlusion plays a significant role for limiting the efficiency. To deal with such issue, vehicle’s images with rear information including lane dash line marking have been collected. It enables the proposed approach to detect and classify multilane vehicles. The dash line detection makes it feasible to detect occluded vehicles and their exact location. The camera has been placed in such a way that it takes the rear view of vehicles images on multiple lanes of the highway. To detect or localize the vehicle on image, background subtraction scheme has been applied.

2) Background subtraction: Considering the significance of Gaussian Mixture Model (GMM) algorithm for background subtraction [6] [10], in this paper an enhanced GMM scheme has been employed for background subtraction and vehicle detection. The proposed GMM model is discussed as follows:

a) Enhanced gaussian mixture model based background subtraction: Unlike conventional threshold-based approaches [16], proposed model applies an enhanced GMM scheme for background subtraction. GMM based background subtraction is nothing else but a pixel-based approach. Consider $x$ be a pixel value at certain time instant. A flexible measure to estimate the probability density function (PDF) of $x$ can be the GMM, in which the PDF comprises the sum of Gaussians. With K component densities the PDF of the Gaussian mixture $p(x)$ can be estimated as:

$$p(x) = \sum_{k=1}^{K} w_k N(x; \mu_k, \sigma_k)$$

(1)

Where $w_k$ represents the weight factor, and $N(x; \mu_k, \sigma_k)$ gives the normal density of mean $\mu_k$ and the covariance matrix $\Sigma_k = \sigma_k I$. GMM as suggested in [26] calculates these parameters to obtain the background. Initially, these parameters are initialized with zero, (i.e., $w_k = w_0, \mu_k = \mu_0, \sigma_k = \sigma_0$). In the case of any similarity, i.e., $\|x - \mu_j\| / \sigma_j < \tau$, with $j \in [1,...,K]$ and $\tau(>0)$ as a certain threshold level, the GMM parameters are updated as follows:

$$w_k(t) = (1 - \alpha)w_k(t - 1) + \alpha M_k(t)$$

(2)

$$\mu_k(t) = (1 - \beta)\mu_k(t - 1) + \beta x$$

(3)

$$\sigma_k^2(t) = (1 - \beta)\sigma_k^2(t - 1) + \beta \| (x - \mu_k(t)) \|^2$$

(4)

Where $M_k(t) = 1$ in the case of the matching element $j$ otherwise is considered as 0.

In case of zero similarity or non-matching elements, the component with minimum $w_k$ is re-initialized, i.e., $w_k = w_0, \mu_k = \mu_0, \sigma_k = \sigma_0$. In above equations (2-4), $\alpha$ represents the learning rate, and $\beta$ is obtained as:

$$\beta = \alpha N(x; \mu_k, \sigma_k)$$

(5)

Here, the weight parameter $w_k$ is normalized iteratively so as to increase to $1$. In [26], researchers sorted Gaussians $w_k / \sigma_k$ in decreasing order so as to perform background subtraction. In background subtraction, GMM applies a threshold value $\lambda$ which is used to the cumulative sum of weights so as to obtain the set $[1,...,B]$. Mathematically, background subtraction is performed using equation (6).

$$B = \arg \min_{k} \left( \sum_{k=1}^{K} w_k > \lambda \right)$$

(6)

In this approach, the Gaussians with the maximum $w_k$ and minimum standard deviation $\sigma_k$ represent the background region. In major GMM models $\mu_k$ and $\sigma_k$ are updated with certain constant learning rate $\beta$ [26]. However, it can’t be effective for dynamic application scenarios such as traffic movement, background changes, and varying lighting or illumination conditions. To deal with such issues, a modification was made in [27]. In [27] the learning rate $\beta$ was assigned in an initial learning process that enabled adaptation under dynamic surface change. In real time applications, there can be pixels which might neither be a foreground nor a background object. However, such pixel is classified either as foreground or background. It leads inaccurate vehicle detection. As proposed in [27], increasing $\beta$ might cause extremely high rate pixel feature variations such as illumination that may make the system vulnerable. Similarly, with the square of the difference between mean and the pixel values might lead higher variance, resulting in continuous increase in illuminations till the saturation of Gaussian mixture over entire pixel color range. Observing both these approaches [26][27], it can be found the earlier [26] lacks dealing with dynamic surface variation, while later [27] suffers from pixel saturation caused due to fast variations (in variance). To deal with such issues, in this paper, an adaptive learning rate based enhanced GMM model has been developed that alleviates such degeneracy, especially in variance by introducing an optimal parameter update paradigm. In the proposed approach, the learning rate has been decoupled for $\mu_k$ and $\sigma_k$. Unlike conventional approaches, an adaptive learning rate $\gamma_k(t)$ has been applied for updating $\mu_k$ that comprises a relative probability factor $R_k = N(x; \mu_k, \sigma_k)$ that signifies whether a pixel belongs to the kth Gaussian component or not.
\[ y_k(t) = y_k(t-1) + \frac{K+1}{K} R_k - \frac{1}{K} \sum_{i=1}^{K} R_i \]  

The implementation of the proposed adaptive learning rate can provide fast Gaussian component mean update as suggested in [27]. It can also enable coping up with fast illumination changes that can ensure precise ROI identification and localization. Now, substituting \( y_k \) as \( \beta \) in (3), it can be found that the self-governing update of the variance can avoid pixel saturation; however, a fast update might result into degeneracy situation. To alleviate this issue, a semi-parametric model has been applied for variance calculation that can significantly enable quasi-linear adaptation, particularly in the case of small changes from the mean and a degraded response for significantly higher deviations. To achieve this, a sigmoid function has been derived as follows:

\[ f_{a,b}(x, \mu_k) = a + \frac{b - a}{1 + e^{-\beta \sigma^2_k(t-1)}} \]  

Where, \( E(x, \mu_k) = (x - \mu_k)^2(x - \mu_k) \). Here, \( S \) plays the role of sigmoid slope controller. Now, substituting (8) in (4), the variance update is obtained as

\[ \sigma^2_k(t) = (1 - \rho) \sigma^2_k(t-1) + \rho f_{a,b}(x, \mu_k(t-1)) \]  

Where, \( \eta = 0.6 \) and \( f_{a,b}(x, \mu) \) limits \( \sigma_k \) to the region \( \mathbb{R} \in \left[ \frac{a+b}{2}, b \right] \). Here, the values of \( a \) and \( b \) are selected in such way that \( \mathbb{R} \) spans over one kth of the pixel range. Thus, applying the proposed adaptive learning rate based GMM model, background subtraction has been performed. The evaluation of the proposed scheme revealed that \( \gamma_k(0) = 0.05 \) can give better performance for background subtraction. Once performed background subtraction, a connected component analysis (CCA) mechanism has been implemented so as to remove irrelevant connected pixels or blobs so as to enable accurate ROI localization.

3) Vehicle region localization: To enhance the vehicle region detection, CCA scheme has been applied that considers valid region, size, and location on the image to remove irrelevant components. Here, a hypothesis that the connected region signifies the Gaussian components belonging to the single lane has been taken into consideration. In the proposed approach, CCA has been performed based on the centroid position. To use the lane information, the width of the individual connected components based on the allied lane has been normalized. The normalized width has been used as the width of the connected component region divided by the width of the lane at the centroid of the connected region. Using the normalized width, it becomes flexible to compare the vehicle size at distinct locations. Thus, employing the enhanced GMM and CCA approaches the exact vehicle regions or the ROI have been localized, which is followed by its feature extraction.

B. Feature Extraction

Once estimating the vehicle region or the ROI, features have been extracted to execute further vehicle classification. In a practical scenario, the vehicles of different categories such as sedan, SUV, MPV, van, truck, etc. would have different features. These high differences in features make classification intricate. As depicted below (Fig. 2) the vehicle (a) represents a MPV, (b) taxi, (c) van and (d) is the other commercial vehicles. These vehicles have different shape, size and color and therefore would have different features too. Considering a broad view of classification where these vehicles have to be classified into two categories, passenger and commercial or other types, to distinguish these vehicles correctly would be highly intricate because these vehicles can have same color, size etc. To enable efficient classification there is the need of certain robust image feature extraction and semantic learning paradigm.

In this paper, the deep learning approach has been applied to perform vehicle or ROI feature extraction. Here, a well-known and robust image feature extraction model based on convolutional neural network (CNN) named AlexNet has been applied to extract ROI features. AlexNet is a multilayered DNN that functions based on convolutional neural network concept and works on ImageNet data. Ironically, the direct implementation of AlexNet DNN with generic computing systems and data elements is not feasible; therefore we have applied a parallel DNN model called CaffeNet [28] with AlexNet. It enabled AlexNet function on general purpose computers. The brief of the AlexNet DNN scheme is presented as follows:

1) AlexNet DNN based feature extraction: In this paper, CaffeNet based AlexNet feature extraction has been performed on vehicle dataset LSVC-2012. The developed feature extraction model has been trained over the localized vehicle ROI data. To enable ROI data for feature extraction with multilayered AlexNet DNN, each vehicle region image has been resized to 256 × 256 dimension. As depicted in Fig. 3, AlexNet comprises five CONVOLUTIONAL LAYERS (CONV1-CONV5) and three FULLY CONNECTED LAYERS (FC6-FC8). The initial layer of this model can have general features resembling Gabor information and blob features. On the contrary, the higher layers comprise significant information for classification; therefore in AlexNet (Fig. 3) five CONVOLUTIONAL LAYERS and two FULLY CONNECTED LAYERS (FC6 and FC7) have been applied to extract features at different layers. Here, each convolutional layer comprises multiple kernels where each kernel signifies a 3D filter connected to the outputs of the preceding layer. In case of fully-connected layers (FC6-FC8), the individual layer comprises multiple neurons containing a real positive value.

Fig. 2. Vehicle images with significantly higher different features
The individual neuron is connected to all the neurons of the previous layer. In this paper, features have been obtained at the two fully connected layers, FC6 and FC7. To achieve better performance, 4096-dimensional features have been obtained at the higher layers of the DNN, FC6 and FC7. These extracted features have been presented in terms of a feature vector $F_V = (f_1, f_2, f_3, ..., f_{4096})$ which has been later processed for dimensional reduction and feature selection. Once retrieving the features, the implementation of dimensional reduction schemes can enable swift and accurate vehicle classification. In this work, two-dimensional reduction algorithms, principle component analysis (PCA) and linear Discriminant analysis (LDA) have been applied to perform dimensional reduction and feature selection. Similar to the AlexNet DNN based feature extraction, SIFT approach has been applied to examine relative performance efficacy.

2) SIFT based feature extraction: This is the matter of fact that feature extraction, selection and its mapping plays a significant role to perform classification. The majority of classification systems are still insignificant because of lower inter-class scatter, particularly with vehicle’s multiclass classification. In practice, the vehicle region or ROI in the image might be very small in size than the overall image size and even the change in lighting can introduce additional intricacies and the insignificant feature that eventually might impact classification accuracy. Here an effort has been made to enhance vehicle detection by applying an enhanced GMM background subtraction model. However, considering existing work and suggestions [25], in this paper, SIFT approach has also been applied to extract ROI features. To retrieve SIFT-based features, four directional filtering 128 SIFT feature descriptors of the each image have been obtained, i.e., 128-dimensional vectors. Similar to AlexNet features, SIFT features has been processed for dimensional reduction using PCA and LDA. It has been followed by SVM-based classification. The retrieved vectors have been projected to PCA algorithm for dimensional reduction. In this paper, the first 64 dimensional vectors have been considered and employing 32 Gaussian components distribution; fisher encoding has been done that eventually generates 4096-dimensional feature vector, which is equivalent to the AlexNet-FC6/FC7.

The discussion of the proposed dimensional reduction approach is presented as follows:

\[ \text{Input Image} \]

- 96 kernels → CONV1
- 256 kernels → CONV2
- 384 kernels → CONV3
- 384 kernels → CONV4
- 256 kernels → CONV5

\[ \text{Fully connected layers} \]

- 4096 neurons → FC6
- 4096 neurons → FC7
- 1000 neurons → FC8

\[ \text{Dimensional Reduction and Classification} \]

C. Dimensional Reduction and Classification

As discussed above, in feature extraction AlexNet as well as SIFT feature descriptors retrieved 4096-dimensional features for each image and therefore to achieve computation and time efficient classification, two predominant dimensional reduction and feature selection approaches, PCA and LDA have been applied. A brief of the applied dimensional reduction approaches is given as follows:

1) PRINCIPLE COMPONENT ANALYSIS: In this work, it is intended to classify vehicles in multiple classes. In general, the feature components extracted from PCA algorithm used to be the most expressive features (MEF), while LDA employs the most discriminating features (MDF) function. In PCA-based approach distinct principle component (PCS) is
generated for an individual class. However, despite of retrieving the distance from the average principal component of each class, the PCA vectors have been trained using SVM classifier. Here, radial basis function (RBF) kernel has been applied for SVM training. SVM has been trained to retrieve the largest feasible classification margin that signifies the lowest value of \( w \) in

\[
\frac{1}{2} w^T w + E \sum \varepsilon_i
\]

(10)

Where \( \varepsilon_i \geq 0 \) and \( E \) is the error tolerance level.

To perform classification, the training vectors have been categorized in labeled pairs \( (x_i, y_i) \) where \( x_i \) states the training vector, while the class label of \( x_i \) is given by \( y_i \in \{-1, 1\} \). In classification, the hyperplane groups highest feasible points of the same class on the same side, while increasing the distance of either class from it. To achieve optimal classification accuracy 10-fold cross validation has been performed. To perform testing, a test image data has been processed for PCS estimation which has been followed by its principle component classification using trained SVM.

2) Linear discriminant analysis: As discussed above, PCA-based schemes employ MEFs to perform classification. However, MEFs can’t be the MDFs all the time. On the contrary, LDA can perform automatic feature selection that can enable efficient feature space for further classification. To alleviate the issue of high dimensionality, LDA has been initiated by employing PCA, where all the vehicle region data or ROI irrespective of the class label has been projected onto a single PCS. The dimension of the PCS has been confined by the total training image minus the number of classes. In this model, two distinct metrics have been estimated, intra-class scatter matrix \( I_{ICW} \) and inter-class scatter matrix \( I_{IOS} \). Mathematically these matrices have been estimated as

\[
I_{ICW} = \sum_{i=1}^{C} \sum_{j=1}^{M_i} (y_j - \mu_i)(y_j - \mu_i)^T
\]

(11)

\[
I_{IOS} = \sum_{i=1}^{C} (\mu_i - \mu)(\mu_i - \mu)^T
\]

(12)

Where \( C \) represents the total number of classes, \( \mu_i \) states the average vector of a class \( i \), and \( M_i \) signifies the number of samples within \( i \). Thus, the average of the average vectors is obtained as

\[
\mu = \frac{1}{C} \sum_{i=1}^{C} \mu_i
\]

(13)

LDA approach focuses on maximizing the inter-class scatter while reducing the intra-class scatter by increasing the ratio \( \frac{\text{det}[S_B]}{\text{det}[S_W]} \). The significance of applying this ratio is that in the case of a non-singular \( I_{IOS} \) matrix, the ratio can be increased when the column vectors of the projection matrix \( W \) can be the eigenvectors of \( I_{ICW}^{-1} I_{IOS} \). Here, the projection matrix \( W \) with \( C-1 \) dimension assigns the training data onto a new space, usually called fisher vector. Thus, \( W \) is applied for projecting all the training samples onto the fisher vector. The retrieved feature vector \( F_{VR} = (f_{1R}, f_{2R}, f_{3R}, \ldots, f_{4096R}) \) has been further used for classification.

In the proposed approach, the obtained vectors have been used to form a know discovery-tree that in the later stage has estimated the nearest neighbors during classification.

In addition to the AlexNet DNN based feature extraction, in this research SIFT has been applied for feature selection, which has been further processed for dimensional reduction using PCA and PLA (Fig. 1).

D. Classification

In this paper, a polynomial kernel based support vector machine (SVM) has been applied to perform vehicle classification. The extracted and dimensionally reduced features from LDA and PCA (Table 1) have been projected and mapped for SVM- based classification. To achieve optimal classification accuracy, 10-fold cross validation has been done. The vehicles have been classified into two broad classes, passenger and other, where passenger class contains vehicle types SUV, van, bus, and cars.

### Table I. Dimensional Reduction and Classification Schemes

<table>
<thead>
<tr>
<th>Data Feature</th>
<th>Dimensional Reduction</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet</td>
<td>PCA</td>
<td>SVM</td>
</tr>
<tr>
<td>AlexNet</td>
<td>LDA</td>
<td>SVM</td>
</tr>
<tr>
<td>SIFT-FV</td>
<td>PCA</td>
<td>SVM</td>
</tr>
<tr>
<td>SIFT-FV</td>
<td>LDA</td>
<td>SVM</td>
</tr>
</tbody>
</table>

Thus, the overall research implementation of the presented work is depicted in Fig. 4

![Fig. 4. Implementation Model](image_url)

The performance evaluation of the proposed vehicle detection and classification algorithm has been discussed in following section.

V. RESULTS AND ANALYSIS

The results obtained are discussed in this section. To perform vehicle detection and classification, a total of 400 images of the vehicles with rear information were used for analysis. Among these images 200 images were from the vehicle category sedan, SUV, etc. or passenger category while