Chapter 3 A Robust Vehicle Detection, Tracking and Speed Measurement Model for Intelligent Transport Systems

3.1 Background

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 ITS solution. Although, a number of techniques have been proposed for moving vehicle detection and tracking; however, the optimization needs cannot be ignored. In this thesis, a robust vision based vehicle detection, tracking and speed measurement model has been proposed and 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 developing for moving vehicle detection, which considers intensity, moving pixel orientation etc. for efficient candidate vehicle detection in traffic video. In order 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. Furthermore, a novel feature clustering scheme with heuristic filtering based blob analysis and adaptive bounding box generation makes this proposed model more efficient for vehicle detection and tracking. In addition, a vehicle speed estimation model has been developed that can play vital role for ITS applications, especially for administration and criminal events tracking purposes.
3.2 Vision Based Vehicle Detection, Tracking and Speed Measurement System

This is the matter of fact that the detection of a moving object (here, vehicle) from traffic video is highly intricate task. Some of the predominant techniques applied to perform moving vehicle detection are temporal differencing approach, optical flow analysis technique, background subtraction based approaches etc. The discussion of these approaches can be found in previous chapters (Chapter-1 and Chapter-2). The typical temporal differencing approach employs two distinct sequential video frames to retrieve background image. However, this approach has certain limitation like it is unable to detect vehicle moving slowly. The approach of optical flow approach can detect object separately by means of camera motion, but unfortunately it is found complicate in computation that makes it less-preferable for real time purposes. On the other hand, in background subtraction approach the complete differences in between the foreground and the background is estimated for each frame to perform moving vehicle detection. However, this approach requires further enhancement to enable efficient and robust moving vehicle detection even with varying lighting conditions, illuminations and background changes. With these motivations, in this research work, an enhanced background subtraction model has been developed for moving vehicle detection. The proposed model considers background subtraction concept for moving vehicle detection but unlike conventional approaches, here numerous algorithmic optimization approaches have been applied 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 the proposed system more robust and efficient. The feature
clustering scheme with heuristic filtering based blob analysis makes this 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 thesis a speed estimation mechanism has been developed, which measures the speed of vehicle in real time movement.

Figure 3.1 Proposed vehicle detection and tracking system

In this section discusses the proposed algorithms and its efficient implementation to achieve intended vision based moving vehicle detection, tracking and speed measurement.
3.2.1 Traffic Video Data Acquisition

In this thesis, to examine the effectiveness of the proposed moving vehicle detection, tracking 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. In addition, some additional standard traffic video data has also been considered for performance evaluation. The input video data are in RGB form, which are further converted into gray color format for processing.

3.2.2 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 research phase, the pre-processing of the video has been done where the input RGB video has been converted into the frame that has been followed by the extraction of the varied constraints including the frame rate, total number of frames, colour, frame size etc. Unlike majority of existing systems, where the initial declaration of the total number of frames is must, in this research work or thesis, an automatic frame and dimensional extraction approach has been incorporated that makes proposed system capable to processing any kind of videos having different features and dimensional characteristics. Once retrieving the frames of the input videos, the RGB images (Figure 3.2) have been converted into gray color image (Figure 3.3), which is then followed by filtering and vehicle segmentation process.
3.2.3 Moving Vehicle Region Detection

Unlike conventional approaches, in this thesis, 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 this work a multilevel optimization model has been proposed that ensures efficient video analysis and
feature mapping for final video tracking purposes. Here, the resulting mapped feature is a gray-scale image having input images of the same size. In this model, the pixel intensity signifies the probability of vehicle in the current frame.

The overall process of moving vehicle detection is discussed as follows:

### 3.2.3.1 Background Extraction

This is the matter of fact that the core of the background subtraction approach is to retrieve the background region of the moving video data. In traffic surveillance system, while recording video on highway; it becomes highly intricate to identify image without any moving vehicle. In order to retrieve such image here implemented background subtraction model. The proposed model applies the mean of all video frame’s pixel values. Thus, retrieving the background image, the Region of Interest (ROI) extraction has been performed. Here, vehicle moving towards camera is tracked where a single lane region is taken traffic track over which the vehicle has to be tracked. In other way, the camera is mounted on road side in such manner that only one lane is visible for vehicle detection and tracking. At first the video RGB frames are converted into gray color using MATLAB function \texttt{rgb2gray}. In the proposed model, 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 3.4). To perform background subtraction different morphological functions and connected component based scheme has been applied, where all the ROI feature vectors in conjunction with each other (connected component) signifies the vehicle region. Further, the morphological closing and thinning operators have been applied to segment the vehicle region.
In the proposed model, 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 presented in the next sub-section of the presented thesis. The developed model intends to avoid any irrelevant object (i.e., waving trees, road marking, etc.) and its movement causing ambiguity for precise vehicle detection and its tracking. The difference between individual frame and allied background model after multiplying both with extracted ROI is used to perform vehicle detection. The background subtracted frame is given in Figure 3.5.
3.2.3.1.1. **Threshold Estimation**

The proposed system employs a thresholding based segmentation scheme that converts grey scale image to binary image. In fact, the selection of an optimal threshold plays a vital role in assuring optimal image segmentation, especially in thresholding based segmentation. Therefore, in this thesis, 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 (3.1).

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

(3.1)

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

3.2.3.1.2. **Directional filtering**

In order to achieve optimal performance, the magnitude of the second derivative of intensity is applied as the edge strength measure, because it facilitates optimal intensity peak detection that usually characterize vehicle in the current video frame. In this thesis, the edge density of the moving vehicle has been estimated 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, a multi-directional filtering has been proposed that estimates the variance of orientations in four directions 0°, 45°, 90° and 135°. Here, 0° indicates horizontal scan, 90° signifies vertical orientation, and 45° and 135° present diagonal orientation. For simplistic implementation, only horizontal and vertical directional filtering has been applied which has been followed by fusion or the amalgamation of all the directional feature vectors to characterize the detected vehicle. 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 thesis, the convolution concept has been introduced with a compass operator (Sobel operator as depicted in Figure 3.6) that retrieves multi-directional edge
intensity \( E_{\theta=0,90,180,45,135} \) 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.

![Figure 3.6: Compass operator](image)

### 3.2.3.1.3. 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 its 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 these 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 introduced due to the broken edges, a two-stage edge generation scheme has been applied. In first stage the strong vertical edges are obtained as given in equation (3.2).
CHAPTER 3. A ROBUST VEHICLE DETECTION, TRACKING AND SPEED
MEASUREMENT MODEL FOR INTELLIGENT TRANSPORT SYSTEMS

\[ \text{Edge}_{vertical}^{strong} = |E_{90}|_{2} \] (3.2)

where \( E_{90} \) represents the vertical intensity edge image that is nothing else but the 2D convolution result of the original image having 90\(^{\circ} \) kernel, \( | \cdot |_{Z} \) represents an operator which is used to achieve a binary outcome of the vertical edges. As this approach intends to retrieve the strong edges, it can’t be stated to be susceptible to the threshold. In the second stage, it is intended to retrieve the weak vertical as depicted through following equations (3.3-3.5):

\[
\text{dilated} = \text{Dilation}(\text{Edge}_{verticalbw}^{strong})_{1\times3} \\
\text{closed} = \text{Closing}(\text{dilated})_{m\times1} \\
\text{Edge}_{vertical}^{weak} = |E_{90}\times\text{closed} - \text{dilated}|_{Z} \\
\] (3.3-3.5)

In this process, the morphological dilation has been introduced that plays significant role in eliminating the impacts of a little skewed edges and a vertical linear structuring element \( m\times1 \) 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 element \( m \). Here, in this thesis, it has been 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 the proposed model, considering the requirement of an effective and efficient system \( m \) has been assigned as \( \text{asm} = (1/25) \times \text{width}_{frame} \) that 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 (3.6).

\[
\text{Edge}_{verticalbw} = \text{Edge}_{verticalbw}^{strong} + \text{Edge}_{verticalbw}^{weak} \\
\] (3.6)

In the proposed model, a morphological thinning operator has been implemented.
which is succeeded by means of a connected component labelling mechanism. These functions are:

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

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

(3.7) (3.8)

Here, the morphological thinning function raises the widths of the resulting edges by one pixel (i.e., increases edge thickness by one pixel). It is then followed by labelling of the vertical edges by connected component labelling operator. In the proposed model, 8 and 4-pixel connectivity has been applied for labelling of the edges. Performing labelling of the connected components, the individual edge has been uniquely labelled as a single connected component having distinctive component number. Thus, the labelled edge frame has been processed by a length labelling process that intends enables edge pixels presenting respective dimensions (i.e., lengths). Consequently, all the pixels allied to the same edge have been labelled with the same number which is proportional to its dimensional length. As, the higher value in the length labelled vehicle video frame represents a long edge, in this thesis a thresholding scheme has been employed to distinguish short edges (short_{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 thesis, the efforts has been made to reduce the false negatives of missed detection. Here, in addition to the edge intensity and variance of orientation, a low threshold value has been applied 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 3.6 and Figure 3.7. The combined vehicle detected is given in Figure 3.8.

3.2.3.1.4. Feature Mapping

In this thesis work and the 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 the
proposed system, 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 (3.9-3.11).

\[
\text{Candidate}_{\text{Vehicle}} = \text{Dilation}(\text{short}_{\text{Verticalbw}})_{mxm} \tag{3.9}
\]

\[
\text{Refined}_{\text{Vehicle}} = \text{Candidate}_{\text{Vehicle}} \times \sum_{\theta=90,180} E_{\theta} \tag{3.10}
\]

\[
f_{\text{map}}(i,j) = N \left\{ \sum_{m=-c}^{c} \sum_{n=-c}^{c} [\text{Refined}_{\text{Vehicle}}(i+m,j+n) \times \text{weight}(i,j)] \right\} \tag{3.11}
\]

Here, 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 the 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 (3.11), \(f_{\text{map}}\) represents the resulting feature map, and \(N\) represents normalization that performs normalization of the intensity (feature mapped values) in the range of \([0, 255]\). The function \(\text{weight}(i,j)\) has been applied that estimates the weight of pixel \((i,j)\) based on the number of orientations of edges in a video frame. Applying \(\text{weight}(i,j)\) function, the proposed approach discriminates the candidate regions (i.e., moving vehicle) from background regions.

The vertical \((E_{\theta=90})\) and horizontal scanning \((E_{\theta=180})\) outputs are given in Figure 3.7 and 3.8 correspondingly.
The combined output of the fused multidirectional filtered feature vectors is given in Figure 3.9.
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 3.10 and Figure 3.11.
Now, combining vertical and horizontal detected vehicles, find the combined vehicle output, given in Figure 3.11.
CHAPTER 3. A ROBUST VEHICLE DETECTION, TRACKING AND SPEED MEASUREMENT MODEL FOR INTELLIGENT TRANSPORT SYSTEMS

3.2.3.1.5. **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, implementing certain morphological dilation operator the close regions can be connected together while ignoring or isolating those regions located far away. In the developed vehicle detection model, a morphological dilation operator having square structuring element has been used that joints vehicle regions in the retrieved binary regions.

3.2.3.1.6. **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 thesis, heuristic filtering scheme has been applied for blob analysis and unwanted blob removal. The proposed heuristic filtering scheme possesses two constraints, which functions for filtering out the blobs which don’t possess vehicle regions or ROI. The first constraint functions for removing minute (very small region) and non-connected isolated blobs using threshold value. In this work, area of the blob region is considered rather than the absolute value for the individual blob. It enables the proposed system to function for the vehicle of any dimension or any size. In addition, the 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.

3.2.3.1.7. **Boundary Boxes Generation**

Retaining the blobs reflecting vehicle in the running frame, it has been enclosed inside boundary boxes. In the proposed model four pairs of the boundary box coordinates have
been estimated using the highest and the lowest point or 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 even outside the initial boundary, dimensional parameters (i.e., height and width) of the boundary box have been padded by a small amount. To make detection more precise, visible and road condition adaptive, in the proposed approach, the large boxes such as borders, highway dividers etc. have been ignored and an additional adaptive padding has been introduced that makes it more accurate and efficient, especially for better tracking purposes. 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 3.12.

### 3.2.4 Vehicle Tracking

In this thesis, 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 the proposed approach, 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. Here, a track (a section of road area defined by user) 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. The 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.

### 3.2.5 Speed Estimation Scheme

In this thesis, the detected moving vehicle possessing its matching ID has been tracked over video frames. To estimate the total number of frames having same object, has been estimated using following equation:

\[
TotalFramesCovered = \text{frame}_{\text{Last},n} - \text{frame}_{\text{First},0}
\]  

(3.12)

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 crossed away from the defined track region. The total number of frames has been multiplied with the crossing duration of each frame, which is measured from the frame rate estimated from the traffic video. In the 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:

\[
Speed = \frac{Distance}{(TF \times \text{Framerate})}
\]  

(3.13)
As depicted in Figure 3.12, the speeds of the vehicles obtained are 15m/s, 16m/s and 11m/s.

3.3 Summary:

Considering limitations of the existing systems, such as conventional background subtraction, noise and illumination sensitivity, etc., in this thesis phase a novel multi-directional filtering and fusion based background subtraction model was developed that considers intensity, moving pixel orientation etc. for moving vehicle detection. The proposed multi-directional intensity strokes estimation scheme was found to be strengthening the system for better moving vehicle candidate detection and tracking so as to distinguish moving vehicle region from other background images. In addition, the implementation of the enhanced thinning and dilation based morphological process has made proposed system more robust and accurate. Performing moving vehicle detection, feature mapping was done where feature clustering and heuristic filtering approach was incorporated, which made blob analysis more efficient to detect precise candidate vehicle region. Later, the boundary box generation was facilitated precise vehicle tracking. In addition, to the efficient moving vehicle detection and tracking, in this thesis, an efficient vehicle speed estimation scheme has been developed that enables real time vehicle
tracking and its speed measurement. Predominantly, this research phase focused 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.

Considering the significance and robustness of optical flow analysis technique for vehicle detection, in next research phase (Chapter-4), an enhanced moving vehicle detection and tracking system has been developed.