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 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).
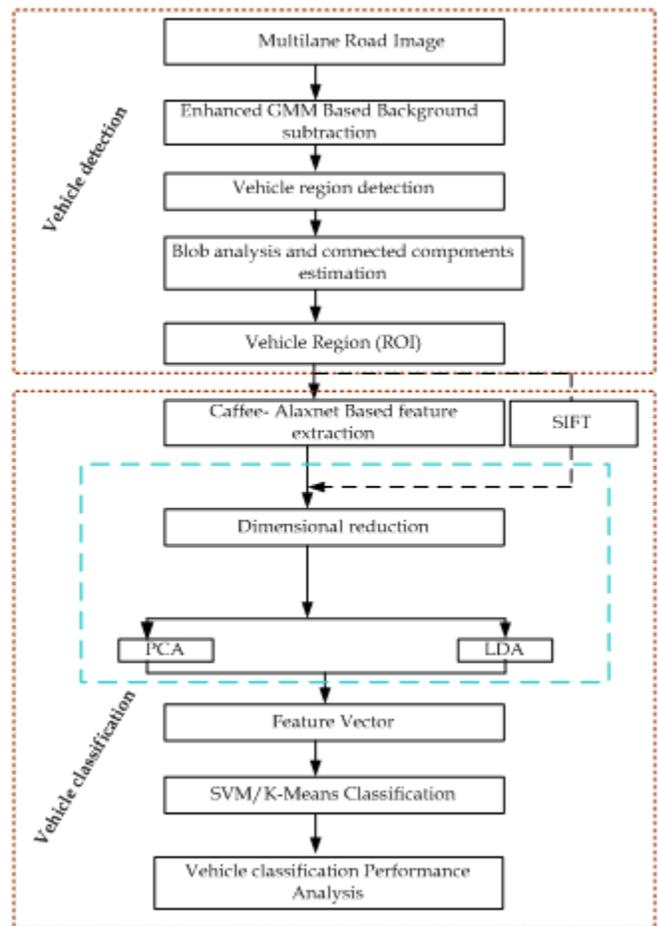

Fig. 1. Proposed vehicle detection and classification system

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

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

Where w_k represents the weight factor, and $N(x; \mu_k, \sigma_k)$ gives the normal density of mean μ_k and the covariance matrix $\sum_k = \sigma_k I$. 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 $j \in [1, \ldots, K]$ and $\tau(> 0)$ as a certain threshold level, the GMM parameters are updated as follows:

$$w_k(t) = (1 - \alpha)w_k(t - 1) + \alpha M_k(t) \quad (2)$$

$$\mu_k(t) = (1 - \beta)\mu_k(t - 1) + \beta x \quad (3)$$

$$\sigma_k^2(t) = (1 - \beta)\sigma_k^2(t - 1) + \beta \|(x - \mu_k(t))\|^2 \quad (4)$$

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), α represents the learning rate, and β is obtained as

$$\beta = \alpha N(x; \mu_k, \sigma_k) \quad (5)$$

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

$$B = arg \min_{K_B} \left(\sum_{k=1}^{K_B \leq K} w_k > \lambda \right) \quad (6)$$

In this approach, the Gaussians with the maximum w_k and minimum standard deviation σ_k represent the background region. In major GMM models μ_k and σ_k are updated with certain constant learning rate β [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 β 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 and classification. As proposed in [27], increasing β 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 μ_k and σ_k. Unlike conventional approaches, an adaptive learning rate $\gamma_k(t)$ has been applied for updating μ_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.

$$\gamma_k(t) = \gamma_k(t-1) + \frac{K+1}{K}R_k - \frac{1}{K}\sum_{i=1}^{K}R_i \qquad (7)$$

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 γ_k as β 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:

$$f_{a,b}(x,\mu_k) = a + \frac{b-a}{1+e^{-SE(x,\mu_k)}} \qquad (8)$$

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

$$\sigma_k^2(t) = (1-\rho)\sigma_k^2(t-1) \qquad (9)$$
$$+ \rho f_{a,b}(x,\mu_k(t-1))$$

Where, $\eta = 0.6$ and $f_{a,b}(x,\mu)_{\mathcal{R}^+}$ limits σ_k to the region $\mathcal{R} \in \left[\frac{a+b}{2}, b\right]$. Here, the values of a and b are selected in such way that $\mathcal{R}$ 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 performing 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 has been 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.
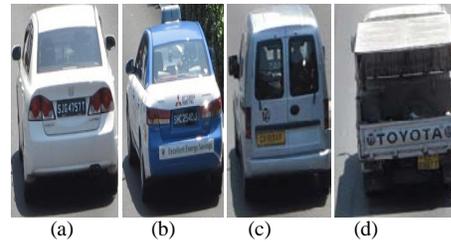

(a)   (b)   (c)   (d)

Fig. 2. Vehicle images with significantly higher different features

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 LSVRC-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.

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 $F_V = (f_1, f_2, f_3, \dots, f_{4096})$ 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:
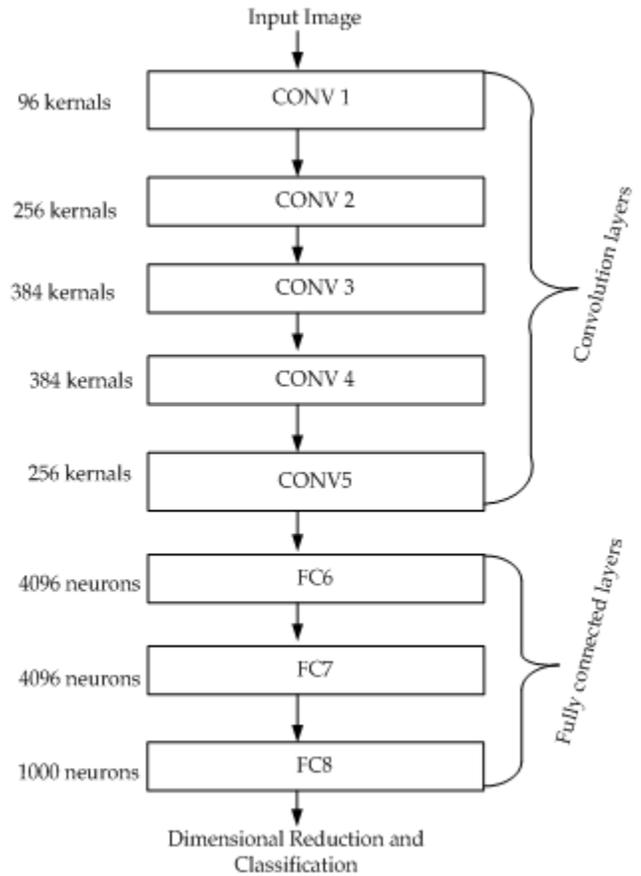

Fig. 3.   AlexNet DNN Architecture

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

$$\frac{1}{2}w^T w + E \sum \varepsilon_i \qquad (10)$$

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

To perform classification, the training vectors have been categorized in labeled pairs $L_i (x_i, y_i)$ 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 matrixes have been estimated as

$$I_{ICW} = \sum_{i=1}^{C} \sum_{j=1}^{M_i} (y_j - \mu_i)(y_j - \mu_i)^T \qquad (11)$$

$$I_{IOS} = \sum_{i=1}^{C} (\mu_i - \mu)(\mu_i - \mu)^T, \qquad (12)$$

Where C represents the total number of classes, μ_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

$$\mu = \frac{1}{C} \sum_{i=1}^{C} \mu_i \qquad (13)$$

LDA approach focuses on maximizing the inter-class scatter while reducing the intra-class scatter by increasing the ratio $\det|S_B|/\det| S_w|$. 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

Data Feature	Dimensional Reduction	Classification
AlexNet	PCA	SVM
AlexNet	LDA	SVM
SIFT-FV	PCA	SVM
SIFT-FV	LDA	SVM

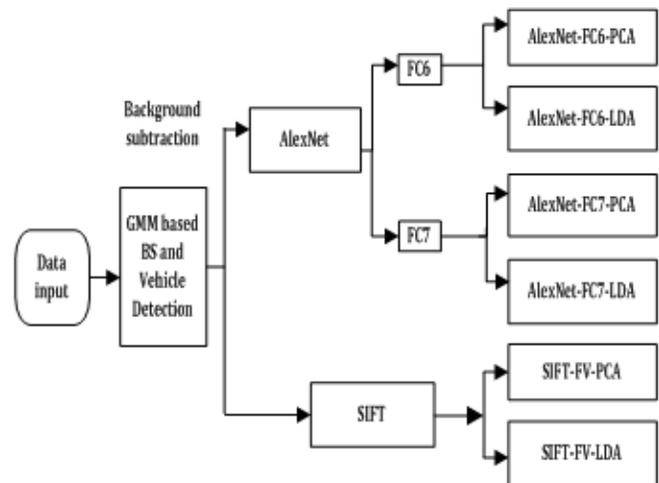Thus, the overall research implementation of the presented work is depicted in Fig. 4

Fig. 4. Implementation Model

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 catagory sedan, SUV, etc. or passenger category while

remaining 200 images were other types including bus, cab, etc. The data was equally selected so as to maintain the similar size of both classes. The initial or the normal size of the images was 4184×3108 that was later resized to 1046×777 for vehicle detection purpose. Once performing background subtraction using Gaussian Mixture Model (GMM), the localized ROI or vehicle region was mapped to the original image with natural resolution. To initiate classification process, the mapped image data was resized to 256×256 size and was fed as input to the AlexNet DNN. With AlexNet DNN based feature extraction, from each image the high layer features FC6 and FC7 were obtained, where each layer possesses 4096 dimensional fetaures. Retrieving the overall features, it was passed to dimensional reduction schemes, PCA and LDA. The dimensionally reduced features were then projected to polynomial kernel function based SVM for multi-class classification. Also, as a parallel model was developed to retrieve features using SIFT scheme. In this approach, the SIFT descriptors of the images were obtained in 128-dimensional vectors which was then projected to PCA for dimensional reduction. The initial 50% of the PCA were selected for analysis, i.e., 64 dimensions. Finally applying 32 Gaussian components distribution, fisher encoding was performed that eventually provided 4096-dimensional feature vector, equivalent to the AlexNet-FC6/FC7. The overall algorithm was developed using MATLAB 2015b tool. Also, VlFeat-0.9.20 toolbox was used to enable swift and easy implementation and processing. The two-class classification results for passenger vehicles and others are presented in Table 2.

TABLE II.   Vehicle Classification Accuracy

	Features	Classification Accuracy (%)
FC-6	GAX6-PCA	96.75
	GAX6-LDA	97.80
FC-7	GAX7-PCA	91.40
	GAX-7LDA	96.30
SIFT	GSIFT-FV-PCA	96.25
	GSIFT-FV-LDA	96.45

[*]GAX signifies proposed GMM detection preceded AlexNet DNN

TABLE III.   Relative Performance Analysis

Techniques	Classification Accuracy (%)
PCA+DFVS[25]	95.85
PCA+DIVS [25]	94.15
PCA+DIVS+DFVS [25]	96.42
SIFT-FV-PCA [29]	92.30
SIFT-FV-LDA [29]	91.30
AlexNet F6-PCA [16]	96.45
AlexNet F6-LDA [16]	97.00
AlexNet F7-PCA [16]	96.80
AlexNet F7LDA [16]	96.10
*GAX6-PCA	96.75
*GAX6-LDA	97.80
*GAX7-PCA	91.40
*GAX-7LDA	96.30
*GSIFT-FV-PCA	96.25
*GSIFT-FV-LDA	96.45

[*]GAX signifies proposed GMM detection preceded AlexNet DNN

Considering comparative performance of the proposed model and others [16], the impact the enhanced GMM for vehicle detection can be easily visualized. The proposed adaptive learning rate based GMM has enabled more accurate vehicle detection under different background and illumination conditions. Now, for classification performance analysis, it can be found that AlexNet FC6 with LDA outperforms other combinations or approaches for two-class classification. Here (Table 2), it can be seen that FC6 feature with LDA dimensional reduction gives classification accuracy of 97.80%, which is higher than FC7 features with PCA (91.40%) and LDA (96.30%). Considering an alternative of DNN, SIFT-based features have also exhibited better with LDA (96.45%) than PCA (96.25%). It states that higher layer DNN features with LDA can give efficient mapped features to perform vehicle classification even with low annotated data.

VI.   CONCLUSION

In this paper, a multilevel optimization measure has been proposed for vehicle detection and classification. Considering the limitations of traditional threshold-based background subtraction schemes, an enhanced adaptive learning rate based GMM algorithm has been developed, which has enabled precise vehicle detection under varying frame background frame features and illumination. To avoid occlusion, in multilane traffic conditions, vehicle's rear features and lane dash markings have been taken into consideration. The application of connected component analysis (CCA) has enabled efficient vehicle region or ROI localization. An enhanced deep convolutional neural network (DNN), named AlexNet has been applied for ROI feature extraction. The implementation of AlexNet-DNN's higher layer features (FC6 and FC7) has exhibited better accuracy, because of higher feature informative contents. As a comparative model, SIFT feature descriptors have been obtained for the ROI. The retrieved 4096-dimensional features from AlexNet-FC6, FC7 and SIFT has been processed for dimensional reduction using PCA and LDA. To perform classification, in this paper polynomial kernel based SVM classifier has been applied that classifies vehicle data into passenger (car, taxi, sedan, SUV) and other types. Results exhibit that AlexNet FC6 features with LDA gives highest classification accuracy of 97.80%, followed by AlexNet-FC6 with PCA (96.75%). The highest accuracy with AlexNet-FC7 has been found lower than AlexNet FC6. Similarly, SIFT features with PCA and LDA (SVM with 10-fold cross validation) has exhibited classification accuracy of 96.25% and 96.45% respectively. The proposed scheme has outperformed other approach because of enhancements introduced regarding adaptive learning rate based GMM. This work has exhibited that adaptive learning rate based GMM with higher layers DNN features can lead optimal vehicle detection and classification. In general, DNN suffers from weight estimation and learning complexity issues, and hence to make this system more effective and time efficient, in future efforts can be made to enhance DNN learning. Concepts such as, shared weight estimation based CNN learning can also be explored to make the proposed system time efficient. In future,

the efficiency of the proposed system could be examined with multiple camera based real time vehicle detection and classification process.

REFERENCE

[1] S. S. Harsha and A. Koteswara Rao, "A Highly Robust Vehicle Detection, Tracking and Speed Measurement Model for Intelligent Transport Systems", *International Journal of Applied Engineering* Research, Vol. 11, No, 5, pp 3731-3740, 2016.

[2] S. S. Harsha and K. R. Anne, "A Robust and Efficient Optical Flow Analysis Based Vehicle Detection and Tracking System for Intelligent Transport System," *International Journal of Computer Science and Information Security (IJCSIS),*Vol. 14, No. 6, pp 605-13,2016.

[3] "USDOT, USDOT intelligent transportation systems research", 2011, http://www.fhwa.dot.gov/research/, Jan 2013.

[4] D. Ang, et al. , "Video analytics for multi-camera traffic surveillance," second international workshop on computational transportation science, Seattle, WA, USA, pp. 25–30, 2009.

[5] A. Ambardekar, et al., "Efficient vehicle tracking and classification for an automated traffic surveillance system," international conference on signal and image processing, Kailua-Kona, HI, USA, pp. 1–6, 2008.

[6] A. Nurhadiyatna, et al., "Gabor filtering for feature extraction in real time vehicle classification system," 9th International Symposium on Image and Signal Processing and Analysis (ISPA), Zagreb, pp. 19-24, 2015.

[7] T. Ma, et al., "Urban vehicle classification based on linear SVM with efficient vector sparse coding," Information and Automation (ICIA), IEEE International Conference on, Yinchuan, pp. 527-532, 2013.

[8] H. C. Karaimer, et al., "Combining Shape-Based and Gradient-Based Classifiers for Vehicle Classification," 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Las Palmas, pp. 800-805, 2015.

[9] X. Chen, et al., "Vehicle representation and classification of surveillance video based on sparse learning," in China Communications, vol. 11, no. 13, pp. 135-141, 2014.

[10] Y. Liu and K. Wang, "Vehicle classification system based on dynamic Bayesian network," Service Operations and Logistics, and Informatics (SOLI), 2014 IEEE International Conference on, Qingdao, pp. 22-26, 2014.

[11] Y. Peng, et al., "Vehicle classification using sparse coding and spatial pyramid matching," 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), Qingdao, pp. 259-263, 2014.

[12] B. Zhang, et al., "Vehicle Classification with Confidence by Classified Vector Quantization," in IEEE Intelligent Transportation Systems Magazine, vol. 5, no. 3, pp. 8-20, 2013.

[13] Y. Hu et al., "Algorithm for vision-based vehicle detection and classification," 2013 IEEE International Conference on Robotics and Biomimetics (ROBIO), Shenzhen, pp. 568-572, 2013.

[14] W. Ma, et al., "Vehicle classification based on multi-feature fusion," Wireless, Mobile and Multimedia Networks (ICWMMN 2013), 5th IET International Conference on, Beijing, pp. 215-219, 2013.

[15] H. Qian, et al., "Vehicle classification based on the fusion of deep network features and traditional features," Advanced Computational Intelligence (ICACI), 2015 Seventh International Conference on, Wuyi, pp. 257-262, 2015.

[16] Y. Zhou, et al., "Vehicle Classification using Transferable Deep Neural Network Features," arXiv:1601.01145v1 [cs.CV] 6 Jan 2016.

[17] Z. Dong, et al., "Vehicle Type Classification Using Unsupervised Convolutional Neural Network," Pattern Recognition (ICPR), 2014, 22nd International Conference on, Stockholm, pp. 172-177, 2014.

[18] H. Gu, et al., "Vehicle size classification for real time intelligent transportation system," Conference Anthology, IEEE, China, pp. 1-5, 2013.

[19] B. Zhang, "Reliable Classification of Vehicle Types Based on Cascade Classifier Ensembles," in IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 1, pp. 322-332, March 2013.

[20] C. Y. Chen, et al., "Vehicle classification and counting system," Audio, Language and Image Processing (ICALIP), 2014 International Conference on, Shanghai, pp. 485-490, 2014.

[21] R. H. Peña-González and M. A. Nuño-Maganda, "Computer vision based real-time vehicle tracking and classification system," 2014 IEEE 57th International Midwest Symposium on Circuits and Systems (MWSCAS), College Station, TX, pp. 679-682, 2014.

[22] Y. Jia, et al., "Caffe: Convolutional architecture for fast feature embedding," in Proceedings of the ACM International Conference on Multimedia. ACM, pp. 675–678, 2014.

[23] A. Krizhevsky, et al., "ImageNet classification with deep convolutional neural networks," in Advances in neural information processing systems, pp. 1097–1105, 2012.

[24] A. S. Razavian, et al., "CNN features off-the-shelf: an astounding baseline for recognition," arXiv preprint arXiv:1403-6382, 2014.

[25] A. Ambardekar, et al., "Vehicle classification framework: a comparative study," EURASIP Journal on Image and Video Processing, vol. 2014, no. 1, p. 29, 2014.

[26] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in Proc. of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 246–252, 1999.

[27] D. S. Lee, "Effective gaussian mixture learning for video background subtraction," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 27, no. 5, pp. 827–832, 2005.

[28] Y. Jia, et al., "Caffe: Convolutional architecture for fast feature embedding," in Proceedings of the ACM International Conference on Multimedia. ACM, pp. 675–678, 2014.

[29] J. S´anchez, et al., "Image classification with the fisher vector: Theory and practice," International journal of computer vision, vol. 105, no. 3, pp. 222–245, 2013.