DUJASE Vol. 4(2) 91-101, 2018 (January & July)

Real Time Feature Based Vehicle Detection and Classification from On-Road Videos

Md. Shamim Reza Sajib and Saifuddin Md. Tareeq

Department of Computer Science and Engineering, University of Dhaka, Dhaka 1000, Bangladesh

E-mail: sajib1717@gmail.com, smtareeq@cse.du.ac.bd

Received on 29.05.18. Accepted for publication on 04.02.19

ABSTRACT

Vision Based vehicle detection and classification has become an active area of research for intelligent transportation system. But this task is very difficult and challenging due to the dynamic condition of roads. In the proposed method, a feature based cost effective detection and classification method is proposed that is suitable for real time applications, provide satisfactory accuracy and computationally cheap. The proposed method uses haar-like image features and AdaBoost classifier for detection. To reduce false positive rate, we propose to use two virtual detection lines (VDL). In order to predict the class of a vehicle, we propose a two level classifier where first classifier separates bigger (bus, truck) vehicles from the smaller one (car, CNG, rickshaw) based on some shape information of vehicles. For the second classifier, we propose to use bag of features (BOF) model which uses the feature efficiently and generates bag of visual words (BOVW). Shape based features are used for first classifier and texture based feature (SURF) is used for second classifier. Error correcting output code (ECOC) framework is used to achieve multi class prediction with SVM to predict the class. Extensive experiments have been carried out on different local traffic data of varying environments to evaluate the detection and classification performance of the proposed method. Experimental results demonstrate that the proposed two level classifier achieves a significant improvement in classification of heterogeneous vehicles in terms of accuracy with a considerable execution time as compared to other methods.

Keywords: Detection and classification of vehicles, Virtual detection line (VDL), Bag of visual words (BOVW), Speeded up robust feature (SURF), Error correcting output code (ECOC), Support vector machine (SVM)

1. Introduction

Vehicle Detection and classification have become an active area of research interest over the past decade. This is a potential area of research as it has some significant applications. Counting vehicles from a video of a particular road is important as there can be rush on a particular time on the road. In Bangladesh traffic jam is very common an unpredictable. Due to limitation of man power, authorities are unable to find reason and control this. Counting vehicles from on road video can help them a lot. They can have the information of traffic flow on a particular road on a particular time, measure rush and reduce the problem. On line traffic surveillance is also becoming common in many countries. Classification or recognition of vehicles also has some important applications like automated toll collections.

Many approaches have been introduced for vehicle detection. Some available methods uses lidar, radar and computer vision. As camera is cheaper than radar or lidar, vision based vehicle detection and classification has become more popular than lidar or radar based detection system. Though computational power has increased dramatically, vehicle detection and classification is not an easy task. The problem is the dynamic environment of the road. The condition of the road cannot be predicted. There can be many human made infrastructures, pedestrians which makes this task a difficult one. Also there are change in background, illusion and heterogeneity of vehicles. Though advanced hardware has become available, real time vehicle detection and classification is still a difficult task. To increase accuracy, researchers have tried deep model which are computationally expensive and not suitable for real time application. The models that need lower computation provide unsatisfactory accuracy. Therefore, the main challenge is to design a model that needs lower computation but provides a satisfactory accuracy and at the same time suitable for real time application.

The target of this paper is to come up with a system which will work in real time. The contributions of the proposed system are reducing partial detection, decreasing false positive in detection and increasing accuracy in classification.

2. Related Work

Background subtraction based method [1] for detection and length based classification which identifies only the long and small vehicles. These two types do not provide much information for designing traffic model. Deep neural network based detection and classification model [2] is very expensive in terms of computational resources and time and is not suitable in real time. Niluthpal *et. al.* [3] generates time spatial images from video frames and gain a very good speed and accuracy in detection. For classification they use two steps KNN classifier on some statistical features from images and gain a satisfactory result. To learn more about related works we divide our approach in two parts, detection and classification. Vehicle detection approaches can be divided into two broad categories: appearance based and motion based methods.

Camera placement plays a significant role in video based vehicle detection. Camera can be moving or static. As occlusion is the main problem in vision based detection system, camera should be placed in some position that minimizes the probability of occlusion. Camera placement depends on the appearance of the vehicles. In [3] camera is placed in an over bridge for taking both incoming and outgoing vehicles. It takes both front and rear view of vehicles. Broggi *et al.* [4] placed a camera to capture the side view of vehicles. Sivram and Trevedi [5] mounted camera in front of the moving vehicles that capture the rear view of front side vehicles. For static camera good choice is a higher position than the level of vehicles that reduces the chances of occlusion. This decreases the chance of partial occlusion caused by vehicles but vehicles appearance changes from first lane to third lane or fourth lane. Yong Tang [6] placed the camera in a high position like over bridge and capture the front view of the vehicles

To detect vehicles from a frame, features are extracted from the frames. Many types of features have been introduced for vehicle detection so far. Sivram and Trevedi [7] used edge features to highlight the side of a vehicle and cast shadow. In recent years simple features like edges or corners are replaced by some strong features due to the robustness and reliability. These features are commonly used in computer vision for both detection and classification. Histogram of oriented gradient (HOG) was extremely well represented for vehicle detection as well as object detection. Inspired by human detection method of Dalal and Triggs [8] Teoh and Brunl [9] has used HOG features nicely to detect vehicles. At first they compute gradients from the images and extract HOG features. The extracted features are trained by a linear SVM classifier. Though HOG features provide a very good detection rate, the main drawback is its calculation speed. As classification will be done after detection, its speed should be good. Haar-like features have been used nicely for face detection. Haar-like features are calculated with the help of integral image. Integral image can be calculated at a very fast speed. Due to its calculation speed and successful use in face detection, it has been also used for vehicle detection in [10] and [11] successfully. Scale invariant feature transform (SIFT) [12] was used in [13] to detect rear faces of vehicles. Though this feature cannot provide better performance than HOG or Haar, it is considerably good in case of occlusion. Lin *et al.* [14] used a combination of SURF [15] and edge features to detect vehicles in the blind spot.

Support vector machine (SVM) is a strong binary classifier. It has been widely used for vehicle detection. The combination of HOG features and SVM classifier have been used for vehicle detection in [9] and [16].

AdaBoost [17] is also widely used in real time vehicle detection. As the classification speed of AdaBoost classifiers is high, it has become popular in real time classification applications. A combination of Haar-like features and AdaBoost classification is used in [18] and [19] for vehicle's rear face detection and perform very good in real time. The purpose of the AdaBoost algorithm is to use the feature to discover the best weak classifiers to form a strong classifier, and has shown its capability to improve the performance of various detection and classification applications. Actually the strong classifier is an ensemble classifier composed of many weak classifiers that just better than a random guess. Tang *et al.* [6] also used this method successfully. This method is very fast and provides high accuracy. But the main drawback of this method is high false positive rate. The method proposed in [18] and [19] achieve a accuracy of 98% with 3%-5% false positive rate which is not tolerable in these type of applications. Yong Tang [6] used this method and achieved good accuracy but had a false positive rate of 3%. This method is very fast and applicable in real time. But false positive rate can cause inapplicable in some sectors.

In recent year deep neural network and model based classification is being used to detect 3D vehicles. Both need high computational resources and execution time. Some motion based approaches have been done by researchers. In [4] an adaptive background model was constructed, with vehicles detected based on motion that differentiated them from the background. Adaptive background modeling was also used in [21], specifically to model the area where overtaking vehicles tend to appear in the camera's field of view. In [3] they used three virtual detection lines and generate time spatial image (TSI) from three frames. The vehicles present in a time spatial image is called TSI object blobs (TOBs). Then canny edges of TOB are generated. After that binary masks of the TOBs are obtained. Then vehicles are detected from multiple TOBs. This method generates a very good result with a good calculation speed and applicable in real time application. But it is not suitable in conditions where there is heavy rush on the road and vehicles are moving in a low speed.

After detecting vehicles in a frame, a pre-trained classifier is used to recognize them. There are two widely used methods for vehicle classification: shape based method, feature based method. Simpson [21] used a multi-dimensional pattern matching algorithm for classification. This method has become successful but need a huge computation time and not applicable in real time. For high computation time with not much satisfactory result, this method is not popular in this area. Petrovic and Coots [22] used edge-based feature for classification but the result was not satisfactory. Zhang [23] proposed ensembles of Gabor feature and PHOG features with an ensembles of some classifiers for vehicle recognition and got a very high accuracy. But this system is very slow but good for image classification. Hsieh *et al.* [24] used SURF features in an excellent way for vehicle classification. They divided the bag of SURF features into multiple grids. Each grid contains features of different classes. They proposed to use an SVM [25] classifier for each grid. The ensemble of that weak classifier makes it strong. Then majority voting scheme is done while testing an image. This method generates a very good accuracy but testing by multiple classifiers makes it slower.

SVM is a strong binary classifier. It provides a very good result for binary classification. For multi class classification, SVM is can be used with some framework like majority voting or error correcting output code (ECOC) [26]. SVM with ECOC has successfully been used in text classification. SVM is used by [15] [27] for vehicle classification. *k* nearest neighborhood kNN and Random Forest (RF) classifiers are also popular in image classification. They both are multi class classifier and generate good result in image classification. kNN is used in [28]#. kNN performs lower when the feature space is imbalanced. If there exists majority of a feature of a class in feature space, accuracy of others classes will be decreased. Random forest provides good result in multi class classification but does not work well in the complex vector space. The tree can grow deep and the size of the forests can be large.

Neural network with single layer is not good for this type of complex feature space. Because a single layer is not enough to separate some classes where each class has a huge variations in features. Neural network with multiple layers [29] also provide good classification result but needs a huge data set and resources also. It is very effective for large scale image classification with millions of data. But it needs a huge resources and time to execute. This method also can be used for vehicle classification to achieve higher accuracy but may be impractical in real life use for its necessity of high resources. The feature based method proposed in [30] does not work satisfactorily when the number of homogeneous classes is big.

3. Proposed Method

Our method is proposed targeting real time applications. Therefore the execution time must be faster in both detection and classification. We are proposing a cost effective and faster model that can easily be implemented for real time applications. Accuracy is a vital issue in these types of application. Therefore we are proposing such a method that is faster as well as provide a very good accuracy.

3.1 Detection

(a) ***Camera placement:*** At the very first stage of detection, we have to select a suitable place to set camera. We propose our method targeting to run on videos taken by some static camera. As occlusion is a big disadvantage in vision based object detection system, camera placement can effect huge on the result. Placing camera on an over-bridge focusing the middle of the road for outgoing vehicles will minimize the chance of occlusion. It will be difficult to find over-bridge everywhere. So camera must be placed in side of the roads where over-bridge not available. The proposed method is designed for both camera positions (fig. 1).

Fig. 1: Left is the view from over bridge and right side shows the view from side

(b) ***Feature extraction*****:** As we propose our method for real time application, we need so select a feature that is fast to compute. In proposed method, haar-like features are chosen for vehicle detection. Haar-like feature is a simple rectangular feature and can be defined as the difference of the sum of pixels of areas inside the rectangle, which can be at any position and scale within the original image. There are three types of rectangle features used in our method: two-rectangle, three-rectangle, four-rectangles (fig. 2). When the position, size or scale of the Haar-like temple are changed, the object feature information, such as the intensity gradient, edge, or contour can be captured. Haar-like features are very useful in real time object detection as it is very fast to compute. Because Haar-like features are calculated with the help of integral images. Integral images are also known as summed-area table.

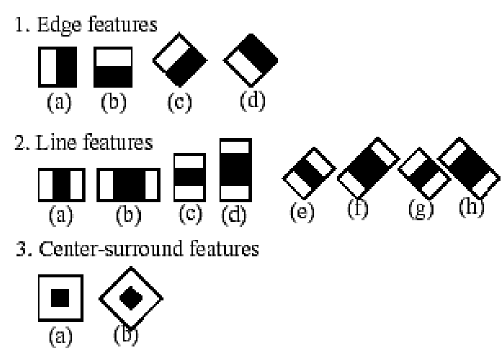

Fig. 2: Haar-like features

Summed-area table is a data structure and algorithm for quickly and efficiently generating the sum of values in a rectangular subset of a grid. The detection is done using Viola-Jones object detection framework [31]. A window of the target size is moved over the input image, and for each subsection of the image the Haar-like feature is calculated. The standard size of the sub window is 24x24. There can be a possible 162,366 possible features. It will be expensive to evaluate all of them. For selecting the best features that separate positive objects from negative, AdaBoost learning algorithm is applied in the proposed method for features selection. In this approach at first it selects the features that present in positive images and not in negative images. Thus the best features are selected which is significant to separate positive from negative.

(c) ***Classifier:*** SVM is faster in training stage than AdaBoost classifiers. SVM becomes quite slower in test stage. We need our method to perform faster in testing.

Some weak classifiers are trained in AdaBoost learning algorithm and then combine them to make a strong classifier. A cascade of 25 classifiers is made in proposed method. The first two classifiers are the strongest one because it is made from the strongest feature selected in AdaBoost feature selection process. These two classifiers can produce a 100% detection rate with a 50% false positive rate. As the number of classifier increases, the false positive rate becomes lower. For training the classifiers, we take 5000 images as positive that contains vehicles and 2000 negative images.

After training we get a cascade of 25 classifiers. In testing stage a sub window of size 40x40 is moved over the input image of size 640x360. The sub-windows that becomes positive after classifying by the first classifier is set as input for the second classifier (fig. 3). The negative sub-windows are discarded in each step. Sub-window that remains positive after going through some steps are said to be positive and the area covering by some joined sub-window is detected as vehicles.

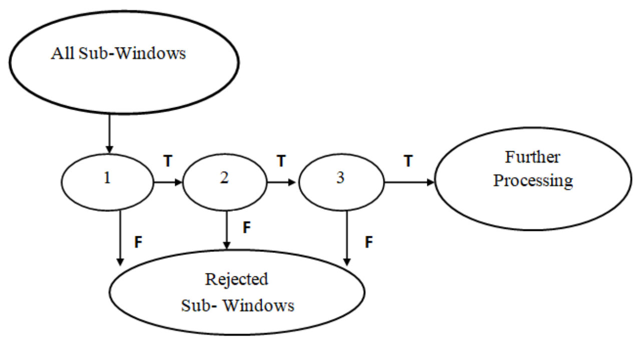

Fig. 3: Testing process of AdaBoost Classifiers

When vehicles move far from camera, the chance of occlusion increases. To minimize this problem, we assume two virtual detection lines (VDL) on the focusing region of camera (fig. 4). This VDL also decreases the chance of false positive rate. Two main reasons of false detection is a) Multiple vehicle occlusion and b) partially appeared vehicles.

Fig. 4: Two virtual detection lines of a frame. Arrow indicates the flow of traffic

Proposed VDL can minimize the chance of multiple vehicle occlusions and can remove partially detected vehicles. The classifier only computes the region of the virtual detection line. So when vehicles appear on the region within virtual detection lines, only then they are detected. When a vehicle appears in the VDL its full rear view must be seen and for this no chance of false detection.

Fig. 5: Partially Appeared Vehicle

As a vehicle must appear within the region of VDL, there will be no possibility to miss any vehicle. By this system, false positive can be reduced to 2% (fig. 5). As we don't need to compute the whole frame and only need to compute VDL regions, this process gives a very good detection rate with a fast computation speed (fig. 6). After detecting the vehicles from a frame, the detected regions of the frame are sent to another classifier which predicts the class of each region.

Fig 6: Detected Vehicles

3.2 Classification

Classification of vehicles is more difficult task because of the intra class difference and inters class similarity. The proposed method considers five categories of vehicles: bus, truck, car, CNG auto-rickshaw and rickshaw. These five types of vehicles are very common in Bangladesh. As mentioned before, shape based classification method needs more computation time than feature based methods. One popular and widely used feature based method is bag of features which is often called bag of visual words (BOVW). The size of the bus and truck is always bigger than the others three. So, in this paper we propose to use a two level classifier where first level separates bigger vehicles from smaller vehicles and then they are classified by a pre-trained classifier used by [30] Using two level classifier, accuracy increases without consuming extra execution time.

(d) ***Feature Selection and Extraction:*** As we propose to use two classifiers, we need two different types of features. The first classifier considers shape features like size, area etc. and the second classifier consider texture features. The shape based features that are used for classification are width which is found from the detected area by proposed detection method. The others features are length, width-length ratio, area which is the actual number of pixels present in the detected vehicle and size which is the product of length and width. These features can be extracted within a negligible amount of time.

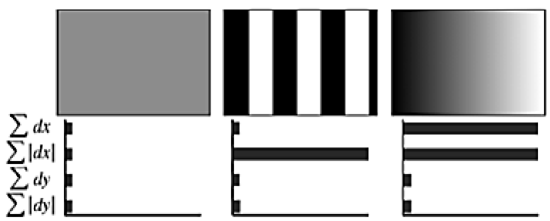

Fig. 7: Left: In case of a homogeneous region, all values are relatively low. Middle: In presence of frequencies in x direction, the value of |dx| is high, but all others remain low. Right: If the intensity is gradually increasing in x direction, both values dx and |dx| are high.

As texture feature we are proposing to use speeded up robust feature (SURF). This feature is robust, scale invariant and less sensitive to varying environments. It uses Gaussian filter for extracting features from a region of image. It finds some key points from an image. SURF's feature descriptor describes a distribution of Haar-wavelet responses within the interest point neighborhood. The SURF descriptor describes an interest area with size 20s. The interest area is divided into 4×4 subareas that is described by the values of a wavelet response in the x and y directions. The wavelet response in the x and y direction is referred to as dx and dy respectively (fig. 7). For each 4×4 sub regions v is calculated where

$$v = \sum dx, \sum |dx|, \sum dy, \sum |dy|$$

Each sub regions need four attributes to represent haar wavelet response, thus the feature vector length of SURF is 64.

Selecting features is not the main fact for good classification. The significant part is how we represent those features. Representing and using those features in a smart way can achieve a good classification model. Therefore, in the proposed method we design a model that represents and uses features efficiently extracted from the images of different class.

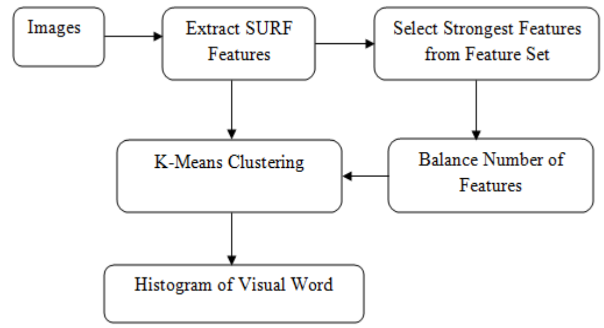

Fig. 8: Block Diagram of Generating Bag of Visual Word

(e) ***Classifier:*** After generating bag of words, feature vector for each class is generated. Then the feature vector is trained through a classifier to predict from four different classes.

As stated above, we use two different classifiers for classification. For first classifier, we take the shape based features from the images and train it by a linear SVM classifier (fig. 9). As the first classifier separates only two classes where the classes has a significant difference in shape, a linear SVM is used as the first classifier. It decides which vehicles are bigger and which are smaller.

For the second level classifier, texture features (SURF) are extracted from the images as the shape based features are not strong enough for separating bus from truck or cng from rickshaw.

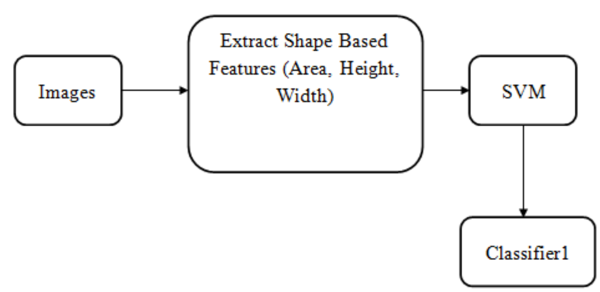

Fig. 9: Block Diagram of Creating Classifier 1

Then the features are clustered using K-Means clustering. After clustering the same type of features should be in the same cluster. Selecting the number of cluster can affect a lot in classification result. Large number of cluster will arise difficulty in finding common features. Small number of clusters makes it difficult to separate features among the classes. Proposed method make 500 clusters from the feature set. BOVW is then generated by computing feature vector of an image by assigning each SURF of the image to one of the K clusters. The histogram of each class is computed in this way (fig. 8). As the second classifier works on multiple classes, a multi class classifier is needed. In proposed method, linear SVM classifier is used with error correcting output code (ECOC) framework (fig. 10). This approach can solve multi class problem for SVM. ECOC framework provides better accuracy than traditional majority voting system. In this framework some linear classifier is trained. Proposed method uses one vs one classification model. In this model there will be six binary learners. In training period each learner assigns some code to each class. When a learner is trained for bus vs car, it assigns a *1* to the code of bus and -*1* to the code of car. Rest two classes get *0* in their codeword for learner *1*. Thus each class has a code of length *6*. When the class of an image is calculated, each class calculates posterior probability with the help of the code. After calculation, the class has the maximum value will be the class of the vehicle (fig. 11).

kNN classifier with ECOC framework also tried in proposed method. When the number of neighbors is set *4* or *5*, it produces the best result. Applying kNN classifier, we have found that both classifiers provide almost same result. So, selecting a classifier is not a significant part in this system, representing features efficiently is the main factor determining the class of a vehicle.

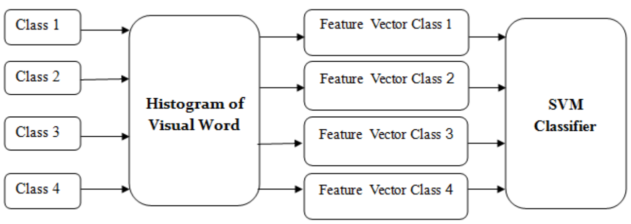

Fig. 10: Block Diagram of Creating Classifier 2

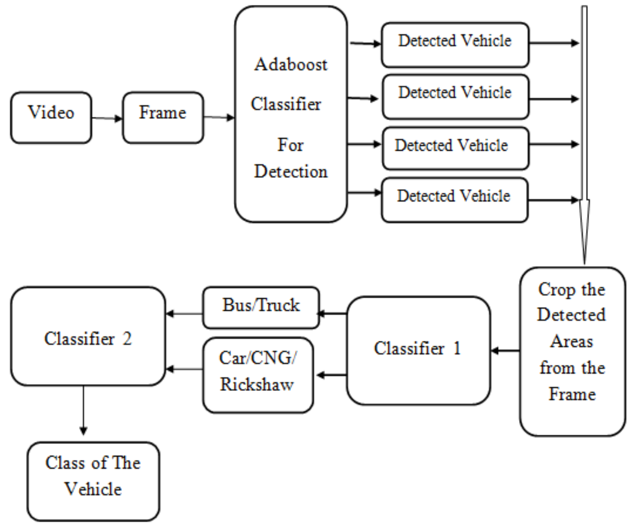

Fig. 11: Block Diagram of the Proposed Method

4. Experimental Result

Multiple experiments have been done to generate the result of the method (fig. 12 & 13). All the experiments have been done on two different data set. Two other methods have been implemented and experimented on the data set to compare the result with the proposed method's result. All experiments have been done on same environment to make the comparison reliable.

Fig. 12: Detected vehicles are classified in a video (Scene 1)

Fig. 13: Detected vehicles are classified in a video (Scene 2)

(a) Data Set Collection: There are 15 videos taken from Kalshi Road, Mirpur, Dhaka, Bangladesh from the side of the roads. Videos are taken on different environments including sunny, cloudy weather (table 1). Some videos are taken in and opposite direction of sunlight. Another 10 videos taken by MVDL [3] authors are used to train and test. Images are extracted from those videos to train. 526 images of cars from Caltech car data set with 2000 images of local Bangladeshi vehicles including bus, car, CNG and rickshaw are trained for detection. For training we take two instances of each image thus the positive image set for detection contains almost 5000 images of vehicles. For negative image set, we use 1300 negative images of Caltech car data set and 700 images of local roads. The negative image set contains almost 2000 images. Those images are resized and used for training. After that many videos are tested and detected vehicles are cropped and saved for classification. We generate two different data sets for training and testing for determining classification results. We add some images of truck with the dataset used for [30].

Data set 1 contains images extracted from videos taken at Kalshi Road, Mirpur. We generate 320 images of each five class for training from 10 videos taken from Klashi Road, Mirpur, Dhaka on different environments (table 2, fig. 14). For testing, images generated from 4 different videos that are not used in training. Test image set contains 53 images of each class.

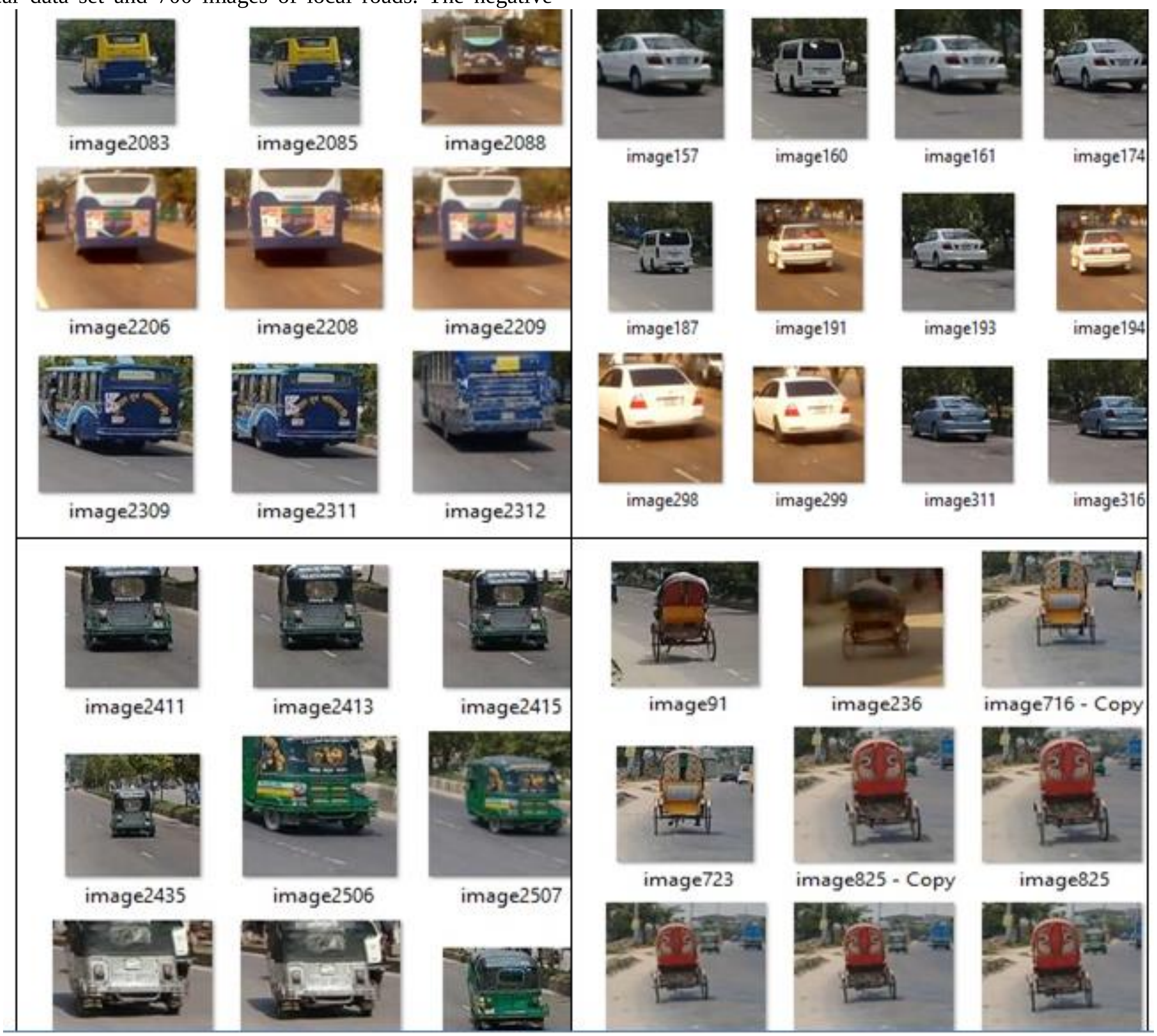

Fig. 14: Image Set for Classification (Data Set 1)

Another data set (Data Set 2) is generated for training and testing. This data set is generated from the videos taken from over bridge of Shahbagh and BUET taken by the authors of [3] method (table 3, fig. 15). For training images, 260 images of each class is taken from 9 videos of different time length. Test data set contains 60 images of each class where the images are cropped from 4 different videos.

Fig. 15: Image Set for Classification (Data Set 2)

Table 1: Collected Data Set Description

Place	No. of Videos	Position	Time	Duration	Environ ment
Kalshi Road, Mirpur, Dhaka	4	Side	9.00 AM	6.36	Sunny
Kalshi Road, Mirpur, Dhaka	4	Side	12.00 PM	6.23	Cloudy
Kalshi Road, Mirpur, Dhaka	5	Side	4.00 PM	9.25	Sunny
Kalshi Road, Mirpur, Dhaka	2	Side	2.00 PM	5.17	Sunny
Cantonment Fly Over Road, Dhaka	3	Side	11.00 AM	3.11	Against Sunlight
Shahbagh, Dhaka	5	Over Bridge	10.0 AM	22.54	Sunny
Shahbagh, Dhaka	3	Over Bridge	2.0 PM	12.21	Cloudy
Shahbagh, Dhaka	2	Over Bridge	4.0 PM	9.46	Partially Sunny
BUET, Dhaka	2	Over Bridge	10.00 AM	8.35	Sunny
BUET, Dhaka	2	Over Bridge	12.00 PM	4.49	Cloudy

Table 2: Data Set 1 For Classification

Place	Position	No of Videos	Test/ Train	Images per class
Kalshi Road, Mirpur, Dhaka	Road Side	10	Train	320
Kalshi Road, Mirpur, Dhaka	Road Side	4	Test	42
Cantonment Fly Over Road, Dhaka	Road Side	1	Test	11

Table 3: Data Set 2 For Classification

Place	Position	No of Videos	Test/ Train	Images per class
Shahbagh, Dhaka	Over Bridge	6	Train	200
BUET, Dhaka	Over Bridge	3	Train	60
Shahbagh, Dhaka	Over Bridge	3	Test	50
BUET, Dhaka	Over Bridge	1	Test	10

(b) Result: Extensive experiments have been carried out to generate the result of the proposed method. The result is computed on a laptop computer of Intel(R) Core(TM) i5-3230M CPU @ 2.60GHz, 4.00 GB RAM, 64-bit Microsoft windows operating system. Without considering occlusion, the average detection rate is 97.81%. Average false positive rate is reduced to 1.8% (table 4).

Experiments have also been done on those two different data sets to proof the reliability and dependability of the proposed method.

Table 4: Detection Accuracy (%) Of Proposed Method

Data Set	Detection Accuracy	False Positive Rate
Data Set 1	97.81	1.25
MVDL [3] Data Set 2	97.23	1.67
Zhang [1]	97.37	2.5

Table 5 refers result generated by [30]. The average classification accuracy is almost 92% for data set 1(side view). The average accuracy of videos of side view is lower than that of top-view. Because from side view it is quite difficult to capture full rear views at all time. From the side, vehicle appearance also changes with the lane. For top view (data set 2), 260 images of each class are trained and total 262 images of four classes are tested. The average classification accuracy for top (middle) view is 94.17%. When we add another class (truck) to the existing data set the average accuracy decreases by 1% for the data set 1 and 2% for data set 2 (table 6).

Table 5: Average Recognition Accuracy (%) Of Four Classes By [30]

Data Set	Bus	Car	CNG	Rickshaw	Average
Data Set1	92.45	88.67	90.56	96.22	91.98
Data set 2	93.33	91.67	93.33	98.33	94.17

Table 6 : Average Recognition Accuracy (%) Of Five Classes By [30]

Data Se	Bus	Truck	Car	CNG	Rick Shaw	Ave Rage
Data Set1	90.56	92.45	88.67	88.67	94.33	90.94
Data Set2.	90.00	91.67	91.67	90.00	96.67	92.00

As the features of the truck have some similarity with the rest classes, the system faces difficulty to separate them accurately. The proposed method provides an average accuracy of 93.58% for dataset 1 and 94.33% for dataset 2. The detail result is shown in table 7. The proposed two level classifier separates two major classes. So BOVW based classifier needs to classify among two and three classes which has some unique features. That's why it generates a better result than [30].

Table 7: Average Recognition Accuracy (%) Of Five Classes By Proposed Method

Data Se	Bus	Truck	Car	CNG	Rick Shaw	Ave Rage
Data Set1	94.33	94.33	90.56	92.45	96.22	93.58
Data Set2	93.33	95.00	91.67	98.33	96.67	94.33

To compare with some other notable existing methods, we use data set 2 both for training and testing. The proposed method shows a good improvement in accuracy. MVDL identifies a broad category of seven classes. MVDL has provided an average accuracy of 86.67% for cng auto rickshaw, 86.67% for car, 91.67% for bus, 88.33% for rickshaw and 93.33% for truck. While they have an accuracy of maximum 89.33% in sunny weather, proposed method provides an average accuracy of 94.33% using SVM classifier with ECOC framework (table 8).

Table 8: Average Recognition Accuracy (%) Generated By Mvdl [3] Method

Data Se	Bus	Truck	Car	CNG	Rick Shaw	Ave Rage
Data Set2	91.67	93.33	86.67	86.67	88.33	89.33

From the result we find that there is a significant improvement in classification of car and rickshaw by the proposed method than MVDL. CNG type vehicle is quite sensitive in the proposed method. Their total computation time is analogous to proposed method.

Table 9: Accuracy (%) Of Classes While Using Surf And Hog Features [6] With Svm

Feature	Bus	Truck	Car	CNG	Rick shaw	Ave rage
SURF	85.00	83.33	83.33	86.67	88.33	85.33
HOG	78.33	81.67	83.33	85.00	86.67	83.00

We also implement method proposed in [6]. The classifier is trained for dataset 2. In the method features are extracted and directly used for classification. The method is implemented considering cell size for HOG feature extraction is 4x4. Average features generated by bus are much higher than the rest of the three as bus is the biggest in size among them. If all the features are taken for classification, the feature will be imbalanced and generate very low accuracy. So, we balance the number of features of different class. We find the highest accuracy when 15,000 features are taken from each vehicle of each class. The average accuracy for dataset 1 is 81.67% and 83.00% for dataset 2 (table 9).

Table 10: Average Recognition Accuracy (%) Generated By [23] Method

Data Se	Bus	Truck	Car	CNG	Rick shaw	Ave rage
Data Set2	91.67	90.00	88.33	91.67	93.33	91.00

Another experiment has been carried out to ensure the main cause of lower accuracy. Using BOVW model using SURF feature [30] generates a good accuracy. Without using BOVW model, extracting SURF features from the image and use those features directly for classifier also provides lower accuracy than that of using BOVW model. This method provides an average accuracy of 86%.

Using SURF feature directly generates lower classification accuracy as like using HOG features in method [6]. But, when we use this feature in a smart way, it provides better accuracy. Method used by [23] provides better accuracy than using SURF directly. Their gridding system and ensembles of classifier of each grid generates better result. It provides 91% accuracy for data set 2 (table 10). But this method consumes much time as there is an ensembles of classifier. For identifying a class, it must calculate the result of all classifiers.

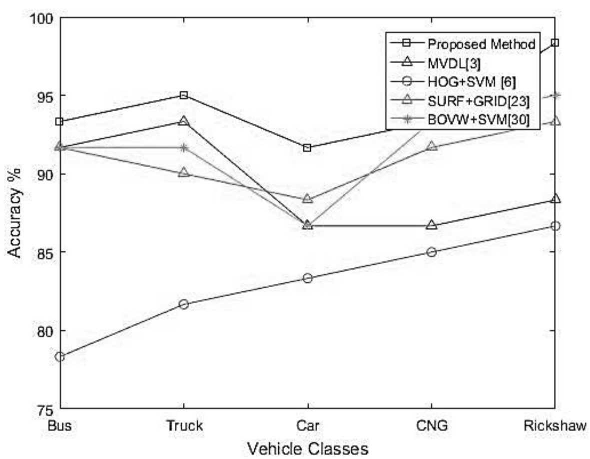

Fig. 15: Comparison of Proposed Method Vs Other Methods

Table 11: Accuracy (%) In Different Illumination

Training Data	Test Data	Average Accuracy
Against Sunlight	To Sunlight	84.77
To Sunlight	Against Sunlight	87.32
Morning	Afternoon	81.23
Afternoon	Morning	79.44

Table 11 illustrates the result of the proposed method in different light. The proposed method is sensitive to different lighting conditions. If we train the classifier for a data set of a light condition and test on another data set of different lighting condition, the average accuracy decreases. Figure 15 demonstrates the comparison of proposed method with some notable current methods. From the graph we clearly understand that proposed method provides a noticeable improvement than the other current methods for heterogeneous vehicles. But proposed method may provide lower accuracy for homogeneous vehicles classification eg. Sedan vs taxi, car vs sedan. It can fail in classifying different brands of cars. But for heterogeneous vehicle classification, this method provides much better accuracy than the others (fig. 15).

Our method needs an average of 99 ms per frame to perform detection and an average of 46 ms per frame to perform classification. The time complexity of the proposed method is $max\ (\ O(n),\ O\ (d),\ O\ (wh\ +\ k\ ))$ where n is the number of features extracted in Viola-Jones algorithm, d the number of input dimensions for SVM classifier, w and h are the width and height and k is the number of extrema found in the input image for calculating SURF.

5. Conclusion

Development in detection and classification of vehicles for real time intelligent transportation system is a very challenging task. The main challenge is, to improve accuracy we need to design a deep model which will be expensive and may not be applicable in real time. If we want to speed up the process, accuracy is expected to decrease. Therefore, we have to design a faster model that will be applicable in real time with a good accuracy. The proposed method uses Haar-like feature based Adaboost classifier that is faster to compute. But this object detection method generates a high false positive rate which is not considerable in many applications. To overcome this problem, the proposed method uses two virtual detection lines (VDL) which reduces the false positive rate. As each and every vehicle must pass through this two VDL, there is not any possibility of missing any vehicles. For classification it uses SURF feature based bag of visual words (BOVW) model to find the class of a vehicle among five classes as the method proposed in [28]. But when we add another vehicle class, its accuracy decreases. Hence, for improving accuracy without consuming much time, a two level classifier is proposed where first level classifies separates bigger (bus and truck) vehicles from the smaller ones (car, cng, rickshaw). Then the second classifier (SURF + BOVW + SVM) identifies the class of a vehicle. Therefore shape based and texture based both features are used to determine the class of a vehicle. The first classifier separates the class based on shape based features while texture based features (SURF) is used by the second classifier to find the exact type of vehicle in the given class. Extensive experiments have been carried out on a number of videos taken from top and side of the road under different environments and different illumination conditions. The proposed feature based method performs better in both detection and classification compared to other existing real time methods with a comparable computation load. Hence, the proposed method can be highly effective in designing video based real time intelligent transportation system.

References

1. Guohui Zhang, Ryan P. Avery, Yinhai Wang, “A Video-based Vehicle Detectionand Classification System for Real-time Traffic Data Collection Using Uncalibrated Video Cameras,”TRB 2007 annual meeting.
2. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.
3. Niluthpol Chowdhury Mithun, Nafi Ur Rashid, S. M. Mahbubur Rahman “Detection and Classication of Vehicles From Video Using Multiple Time-Spatial Images,”IEEE transactions on intelligent trasportation systems, vol. 13, NO. 3, September 2012.
4. A. Broggi, A. Cappalunga, S. Cattani, and P. Zani,“Lateral vehicles detection using monocular high resolution cameras on TerraMax,” in Proc. IEEE Intell. Veh. Symp., pp. 1143-1148, Jun. 2008.
5. Sayanan Sivarama, Mohan Manubhai Trivedi, “Integrated Lane and Vehicle Detection, Localization, and Tracking: A Synergistic Approach,” IEEE transactions on intelligent trasportation systems, vol. 14, no. 2,June 2013.
6. Yong Tang, Congzhe Zhang, Renshu Gu, Peng Li, and Bin Yang. “Vehicle detection and recognition for intelligent traffic surveillance system. Multimedia tools and applications”, 76(4):5817-5832, 2017.
7. Sayanan Sivarama, Mohan Manubhai Trivedi, “Deep Nural Network,” IEEE transactions on intelligent trasportation systems, vol. 14, no. 2,June 2013.
8. Navneet Dalal and Bill Triggs, “Histograms of oriented gradients for human detection,” In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 1, pages 886-893. IEEE, 2005.
9. S. Teoh and T. Brunl, “Symmetry-based monocular vehicle detection system.” Vis. Appl. [Online]. 23(5), pp. 831-842, 2011.
10. W. C. Chang and C.-W. Cho “Real-time side vehicle tracking using parts-based boosting,” IEEE Int. Conf. SMC, pp. 3370-3375, 2008.
11. W.Liu, X.Wen, B.Duan, H.Yuan, and N.Wang, “Rearvehicledetection and tracking for lane change assist,” in Proc. IEEE Intell. Veh. Symp., pp. 252-257, Jun. 2007.
12. D. Lowe “Object recognition from local scale-invariant features,” in Proc. Int. Conf. Comput. Vis., pp. 1150-1157, 1999.
13. X. Zhang, N. Zheng, Y. He, and F. Wang, “Vehicle detection using an extended hidden random field model,” in Proc. 14th Int. IEEE Conf. ITSC, pp. 1555-1559, Oct. 2011.
14. B.-F.Lin, Y.-M.Chan, L.-C.Fu, P.Y.Hsiao, L.-A.Chuang, S.-S.Huang, and M.-F. Lo, “Integrating appearance and edge features for sedan vehicle detection in the blind-spot area,” IEEE Trans. Intell. Transp. Syst., vol. 13, no. 2, pp. 737747, Jun. 2012.
15. Z.Chen, N.Pears, M.Freeman, and J.Austin, “Road vehicle classication using support vector machines,” in Proc. IEEE Int. Conf. Intell. Comput. Intell. Syst., Shanghai, China, 2009, pp. 214218.
16. Sayanan Sivarama, Mohan Manubhai Trivedi, “Active learning for onroad vehicle detection: A comparative study,” Mach. Vis. Appl. Special Issue Car Navigation and Vehicle Systems, pp. 113, Dec. 2011.
17. Yoav Freund, Robert Schapire, and Naoki Abe.“ A short introduction to boosting,” Journal-Japanese Society For Artificial Intelligence, 14(771- 780):1612, 1999.
18. J. Cui, F. Liu, Z. Li, and Z. Jia, “Vehicle localization using a single camera,” in Proc. IEEE IV, Jun. 2010, pp. 871876.
19. D. Withopf and B. Jahne, “Learning algorithm for real-time vehicle tracking,” in Proc. IEEE ITSC, pp. 516-521, Sep. 2006.
20. J. Wang, G. Bebis, and R. Miller, “Overtaking vehicle detection using dynamic and quasi-static background modeling,” in Proc. IEEE CVPR, Jun. 2005.
21. K Simonson, “Multinomial Pattern Matching: A Robust Algorithm for Target Identification,” Automatic Target Recognizer Working Group, Huntsville, 1997.
22. V. Petrovic and T. Cootes, “Analysis of features for rigid structure vehicle type recognition,” in Proc. Brit. Mach. Vis. Conf., pp. 587-596, 2004.
23. Bailing Zhang, “Reliable Classication of Vehicle Types Based on Cascade Classier Ensemble,” IEEE transactions on intelligent transportation system, vol. 14, no. 1, March 2013.
24. Jun-Wei Hsieh, Li-Chih Chen, and Duan-Yu Chen. “Symmetrical surf and its applications to vehicle detection and vehicle make and model recognition,” IEEE Transactions on intelligent transportation systems, 15(1):6-20, 2014.
25. H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, “SURF: Speeded up robust features,” Comput. Vis. Image Understand., vol. 110, no. 3, pp. 346359, 2008.
26. Thomas G Dietterich and Ghulum Bakiri. “Solving multiclass learning problems via error-correcting output codes,” Journal of artificial intelligence research, 2:263-286, 1995.
27. E. Rivlin, M. Rudzsky, M. Goldenberg, U. Bogomolov, and S. Lapchev “A real-time system for classication of moving objects,” Comput. Vis. Image Understand., vol. 110, no. 3, pp. 346359, 2008.
28. E. M. Kornaropoulos and P. Tsakalides, “A novel KNN classier for acoustic vehicle classication based on alpha-stable statistical modeling,” in Proc. IEEE Workshop Stat. Signal Process., Cardiff, U.K., 2009.
29. A. Goyal and B. Verma, “A neural network based approach for the vehicle classication,” in Proc. IEEE Symp. Comput. Intell. Image Signal Process., Honolulu, pp. 226-231, 2007.
30. Md. Shamim Reza Sajib and Saifuddin Md. Tareeq, “A Feature Based Method for Real Time Vehicle Detection and Classification From On-Road Videos,” 20th International Conference of Computer and Information Technology (ICCIT), 2017.
31. Viola, Paul, and Michael Jones. ”Rapid object detection using a boosted cascade of simple features.” Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. IEEE, 2001.