Comparative Analysis of Vehicle Detection in Urban Traffic Environment using Haar Cascaded Classifiers and Blob Statistics

Yumnah Hasan
Electrical Engineering Department
Bahria University
Karachi, Pakistan
yumnahhasan.bukc@bahria.edu.pk

Muhammad Umair Arif
Electronics and Power Engineering Department
NUST-PNEC
Karachi, Pakistan
umair.arif@pnec.nust.edu.pk

Amad Asif
Electrical Engineering Department
Bahria University
Karachi, Pakistan
amad_asif@yahoo.com

Rana Hammad Raza
Electronics and Power Engineering Department
NUST-PNEC
Karachi, Pakistan
hammad@pnec.nust.edu.pk

Abstract*—The applications of computer vision are widely used in traffic monitoring and surveillance. In traffic monitoring, detection of vehicles plays a significant role. Different attributes such as shape, color, size, pose, illumination, shadows, occlusion, background clutter, camera viewing angle, speed of vehicles and environmental conditions pose immense and varying challenges in the detection phase. The native urban datasets namely NIPA and TOLL PLAZA acquired in complex traffic environment are used for research analysis. The selected datasets include varying attributes highlighted above. The NIPA dataset has total of 1516 vehicles whereas the TOLL PLAZA dataset contains 376 vehicles in an entire video sequence. This paper provides comparative analysis and insight on performance of cascade of boosted classifier using Haar features versus statistical analysis using blobs. Haar features help effectively in extracting discernible regions of interest in complex traffic scenes and has minimum false detection rate as compared to blob analysis. The detection results obtained from the trained Haar cascade classifier for NIPA and TOLL PLAZA datasets have 83.7% and 88.3% accuracy respectively. In contrast blob analysis has detection accuracy of only 43.8% for NIPA and 65.7% for TOLL PLAZA datasets.

Keywords—detection; urban; traffic; Haar cascade classifier; blob analysis

I. INTRODUCTION

Vision based surveillance systems typically are becoming more acceptable in a wide range of applications including Intelligent Transportation System (ITS) due to its non-intrusive nature, continuous development of promising algorithms, easy infrastructure setup, cost effectiveness, low maintenance and large information acquisition for analysis. In road monitoring and surveillance, detection of discernible region of interest in the traffic scenes is a fundamental task. There are several techniques developed for the detection of vehicles. However, the research community is still identifying robust and generic approaches to improve the detection in diverse climatic conditions and complex urban traffic scenes. Varying attributes such as object shape, color, size, pose, illumination, shadows, occlusion, background clutter, camera viewing angle and speed of vehicles add an additional layer of complexity to the detection process.

One of the classical yet competitive real-time object detection framework was introduced by Viola and Jones [1]. The approach was primarily designed for face detection application, though is expandable to variety of objects. The authors have utilized an integral of an image to evaluate rectangular (Haar like) features. Later after training phase, the faces are separated from non-face images using cascaded classifiers. This serves as the baseline of our proposed research.

On the other hand, blob analysis is a machine vision based method used for the analysis of consistent image region. Vehicle detection using blob analysis utilizes motion estimation. The moving objects correspond to the group of foreground pixels known as 'blobs'. Morphological operations which include removal of noise are applied on a video sequence to obtain the correct detection results.

In general, there are two types of vehicle detection approaches: appearance based and motion based [13]. Broggi *et al* [3] reported a motion based approach in which motion of differential foreground from background is detected. They have used an adaptive background model for extraction of foreground however, identifying between pedestrian and vehicle is not error free. Jazayeri *et al* [5] utilized a method for separating the background model and moving vehicles in a video sequence by using a hidden Markov Model. According to the joint distribution of horizontal position and velocity of scene, as well as characteristics of scene, they have modeled the image of moving vehicles. Jodoin *et al* [10] introduced a tracker that was able to handle multiple objects of various shape and sizes. They have used background subtraction for

978-1-5090-4171-8/16/$31.00 ©2016 IEEE

moving object detection. For segmentation and occlusion, they have developed a state machine and interface blob. The detection rates of motion based methods are usually high because motions are easily detected without performing any complex computations. However, the error margin of false detection is comparatively high in these methods as they do not take any other features of vehicle into consideration. Subsequently, they are more sensitive to noise.

Appearance based methods are more complex than motion based methods. Their dependencies on the characteristics of scenes are quite high. Pankanti et al [7] from IBM Research have implemented multi-view detection system that relies on a set of motion-let classifier on urban environments including crowded scenes. The same group proposed motion-let approach [8] where clustering is based on motion and motion-let approach using large-scale feature selection implemented on crowded scenes with many occlusions. They have reported false detections due to limited negative images in the training set. Mithun *et al* [9] proposed a detection and classification method based on multiple time-spatial images (TSIs), each obtained from a virtual detection line on the frames of a video. The multiple TSIs are used to increase the accuracy of correct detection. The authors identified the occluded vehicles and reduced the dependency of pixel intensities between still and moving objects. A novel Bayesian fusion algorithm based on Gaussian mixture model was implemented on vehicular traffic in complex challenging environment by Chen and Qin [11]. Their proposed method is computationally expensive and performs low for occluded vehicles. One of the state of the art work is presented in [6] where the authors have implemented automatic detection based on appearance information. The dataset used is acquired under varying illumination conditions. For each set a separate detector was trained for detection. The test set was divided into two different categories namely high crowded scenes and low crowded scenes. They have introduced occlusion handling technique by training a detector on the basis of several occluded images. Occlusion handling is done using poison image reconstruction through gradient. Parallel feature selection over multiple planes including grey scale, red, green and blue channels with standard adaboost detection helps in acquiring more accurate results. Their proposed method produces efficient results in crowded scenes.

In this paper, a comparative analysis and insight on performance of cascade of boosted classifier using Haar features (appearance based) versus statistical analysis using blobs (motion based) are discussed for native complex urban datasets. Details of NIPA and TOLL PLAZA datasets are tabulated in Table 1 and Table II. The Haar cascade classifier has a trained set of positive and negative images in the database for comparison of true and false objects. Performance wise, Haar features provide promising results in the detection process as compared to area characteristics.

Rest of the paper is organized as follows. Section 2 provides dataset details and system settings. Section 3 covers the scheme of implementation. Experimental results and analysis are provided in Section 4. Finally, Section 5 summarizes and concludes the discussion with pointers to future work.

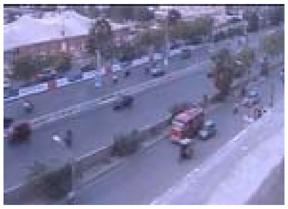
II. System Settings

Native urban datasets named NIPA and TOLL PLAZA [12] are utilized for the research analysis with tabulated details provided in Table I and Table II. The datasets include complex urban traffic scenes where road markings and signs are significantly low. In addition, type of vehicles range from motorbikes to long trucks.

The NIPA dataset consists of three lanes and TOLL PLAZA dataset consists of two lanes with different type of vehicles flowing in alternate directions and varying speeds. In NIPA dataset, cast shadow size is small. Camera is providing low resolution side view of the traffic scene. Due to the side view angle, vehicles pose along with background clutter is also generating considerable complexity. Total of 1516 and 376 vehicles are assessed for NIPA and TOLL PLAZA datasets respectively.

The algorithms were tested on a Core i-5 2.30 GHz, 8 GB RAM platform. OpenCV library is used for training the classifier, VLfeat is used for plotting Receiver Operating Characteristic Curve (ROC) so as to illustrate the performance of a binary classifier [4]. Simulations are done using Matlab® 2014 and Visual Studio 2013. The computational time for training a classifier is around 4-6 hours.

TABLE I. Highlights of the NIPA dataset [12]

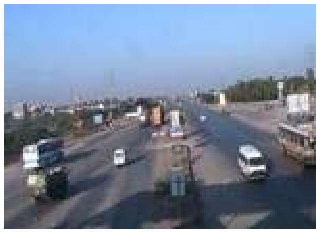
Video Frame	
Sequence Type	Outdoor
Video Length	00:08:16
Image Size	640 x 420
Shadow Strength	Low
Shadow Size	Small
Object Class	Vehicle
Object Size	Small
Object Speed (Pixels)	Slow/Fast
Noise Level	Medium
Camera position	Side view

III. Implementation

For blob statistics, as an initial step morphological operations like hole filling and removal of noise are performed, followed by vehicle detection based on area characteristics. The blob is a connected group of foreground pixels which corresponds to a moving object.

978-1-5090-4171-8/16/$31.00 ©2016 IEEE

TABLE II. HIGHLIGHTS OF THE TOLL DATASET [12]

Video frame	
Sequence Type	Outdoor
Video Length	00:20:00
Image Size	720 x 576
Shadow Strength	High
Shadow Size	Large
Object Class	Vehicle
Object Size	Medium
Object Speed (Pixels)	Slow
Noise Level	Medium
Camera position	Front view

Haar cascade classifier utilizes two different trained classifiers for NIPA dataset. For TOLL PLAZA dataset, one trained Haar cascade classifier is used.. The trained cascade classifier for each dataset is unique on the basis of positive and negative samples and number of stages. The false positive rate can be decreased by increasing the number of stages. In the first case for NIPA dataset the cascade classifier # 1 has 168 positives, 500 negatives and 16-stages, whereas the cascade classifier # 2 has 580 positives, 1500 negatives and 16-stages. In the second case the cascade classifier has 1000 positives, 2000 negatives and 18-stages for the TOLL PLAZA dataset.

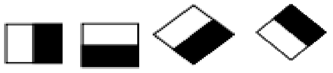
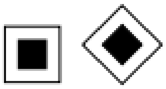
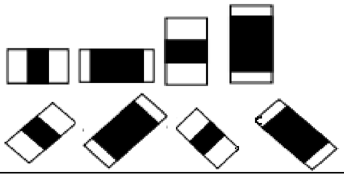
Initially, samples are obtained for each classifier which consists of positive and negative images. The negative samples are random images, which are not required for the detection. However, the positive samples correspond to cropped template of objects of interest.

A. Haar Cascade Classifier

Haar cascade classifier utilizes Haar like features proposed by Paul Viola and Michael Jones [1]. Haar features prototypes [2] were used and are shown in Table III. These 14 features include 2 center-surround features, 4 edge features and 8 line features. A complete and rich set of features is generated by scaling these prototypes independently in vertical and horizontal direction. The cascade function is trained from a set of positive and negative images. The ensuing paragraphs will provide the details regarding the procedures and techniques involved in training a Haar cascade classifier and it's testing on the datasets.

One trained classifier is used for TOLL PLAZA while two different trained classifiers are utilized for NIPA dataset. Trained classifier of TOLL PLAZA produced good results, thereby eliminating need of second classifier. NIPA results obtained from first classifier are less accurate due which second classifier is trained to make the detector sparse and efficient. The cascade function was trained from a set of positive and negative images. The ensuing paragraphs will provide the details regarding the procedures and techniques involved in training a Haar cascade classifier and it's testing on the dataset.

TABLE III. SAMPLES OF USED FEATURES PROTOTYPES [2]

Edge Features	
Center-surround Features	
Line Features	

For both datasets, a set of positive and negative samples are used for training a classifier. Negative images are downloaded from GitHub [14]. A total of 2000 negative images are saved for each datasets, these images are tabulated in a text file for path location and image file extraction. Next, positive samples for both datasets are obtained by cropping a positive sample vector from the image sequence using object marker tool and shown in Fig. 1and Fig. 2.

Thereafter, positive and negative samples are converted into a vector file. The sample image size is significant so as to define number of cascade classifier stages and to maintain adequate resolution that in-turn will be used for feature extraction.

Boosting technique is used to train each classifier by using weighted average of the decision made by decision stumps (one level decision tree). The current location of sliding window (which may be positive or negative) is defined by the labels present in every stage of classifier. Having a negative label indicates zero detection and the classifier moves to the next label. Alternatively, positive label indicates object found and the detected region is fed to the next stage. The detector in the final stage classifies the object with positive label.

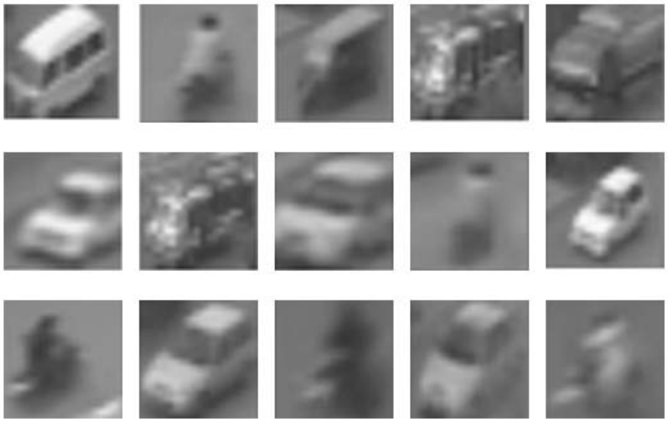


Fig. 1. Positive samples for training a Cascade Classifier [NIPA Dataset]

978-1-5090-4171-8/16/$31.00 ©2016 IEEE

For NIPA dataset, the numbers of stages are 16. Therefore, the false alarm rate for a cascade classifier having 16 stages is approximately 0.5^{16} =1.525 x 10^{-5} and a hit rate of approximately 0.999^{16} = 0.984. The time taken by a cascade classifier to train with 580 positives images, 1500 negatives images and 16 stages on Core i5 is around 4 to 5 hours. While training a cascade classifier for TOLL PLAZA dataset having 2000 negative images and 1000 positive images on the same system configurations required 6 to 7 hours. The false alarm rate is around 0.5^{18} = 3.814 x 10^{-6} and a hit rate of around 0.999^{18} = 0.9821 having 18 stages in actual. Haar training tool is executed to train the classifier in both cases. Finally, cascade classifier was converted into an XML file by using Haar converter tool in C++.

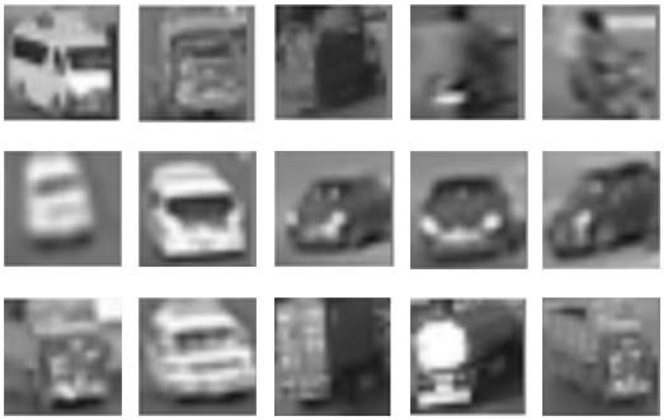


Fig. 2. Positive samples for training a Haar Cascade Classifier [TOLL PLAZA Dataset]

B. Blob Analysis

Blob statistics utilizes connected groups of foreground pixels called ‘blobs’ which correspond to moving objects. An array of tracks representing moving vehicles is generated so as to maintain the state of a tracked vehicle. Noisy detections in a frame tend to result in short-lived tracks. Next foreground detector is used to obtain binary motion segmentation. Subsequently, morphological operations are performed on the resulting binary mask to remove noisy pixels and to fill the holes in the remaining blobs.

The bounding boxes are updated by using the Kalman filter which predicts track centroids in the current frame. Similarly, cost function is minimized to assign object detections in a current frame to existing track. Finally, assigned tracks are identified and lost tracks are eliminated.

IV. Results and Analysis

The vehicle details in each lane of NIPA dataset is given in Table IV. The blob analysis performance is low and detects only 665 vehicles out of 1516. Haar classifier # 1 detects 1242 out of 1516 vehicles. Similarly, Haar classifier # 2 detects 1270 out of 1516 vehicles. Quantitative analysis using these approaches is tabulated in Table IV. Similarly, detection results using blobs analysis and Haar classifier are laid over input images and shown in Fig. 3 and Fig. 4 for visual inspection.

The ground truth of TOLL PLAZA dataset is shown in Table V. The detection results obtained from Haar cascade classifier are better than blob analysis. The numbers of total detected vehicles by using blob analysis are 247 out of 376. However, Haar cascade classifier detected 332 vehicles out of 376. The results of each method are tabulated in Table V. The pictorial results of blob analysis and Haar classifier are shown in Fig. 3 and Fig. 4. The detection results obtained from the trained Haar cascade classifier for NIPA and TOLL PLAZA datasets have 83.7% and 88.3% accuracy respectively. In contrast blob analysis has detection accuracy of 43.8% for NIPA and 65.7% for TOLL PLAZA datasets.

TABLE IV. NIPA Dataset Ground Truth and Qualitative Results Obtained from BLOB Analysis and Haar Cascade Classifier 1 & 2

Nipa				
Category	*Lane1*	*Lane2*	*Lane3*	*Total*
Total	687	422	407	1516
Blob Analysis				
Category	*Lane1*	*Lane2*	*Lane3*	*Total*
Total	216	228	221	665
Correct Vehicle Detection	31.4%	54.0%	54.2%	43.8%
Cascade Classifier # 1				
168 positives, 500 negatives, 16-stages				
Category	*Lane1*	*Lane2*	*Lane3*	*Total*
Total	477	396	369	1242
Correct Vehicle Detection	69.4%	93.8%	90.6%	81.9%
Cascade Classifier # 2				
580 positives, 1500 negatives, 16-stages				
Category	*Lane1*	*Lane2*	*Lane3*	*Total*
Total	507	404	359	1270
Correct Vehicle Detection	73.8%	95.7%	88.2%	83.7%

TABLE V. TOLL PLAZA Dataset Ground Truth And Qualitative Results Obtained from BLOB Analysis and Haar Cascade Classifier

TOLL PLAZA			
Category	*Lane1*	*Lane2*	*Total*
Total	122	254	376
Blob Analysis			
Category	*Lane1*	*Lane2*	*Total*
Total	104	143	247
Correct Vehicle Detection	85.2%	56.3%	65.7%
Cascade Classifier			
1000 positives, 2000 negatives, 18-stages			
Category	*Lane1*	*Lane2*	*Total*
Total	111	221	332
Correct Vehicle Detection	90.9%	87.0%	88.3%

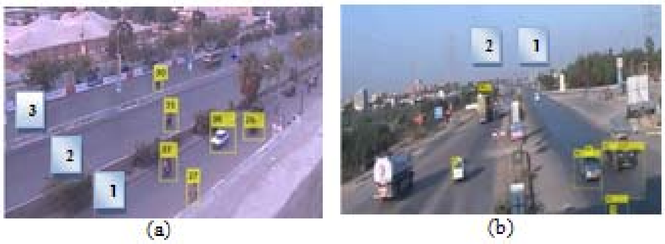


Fig. 3. Detection results obtained from Blob Analysis. Lanes are labeled as 1, 2 and 3 (a) NIPA (b) TOLL PLAZA Datasets

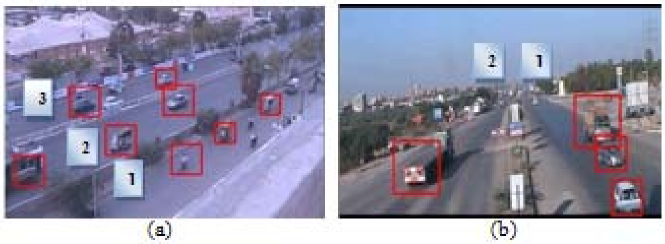
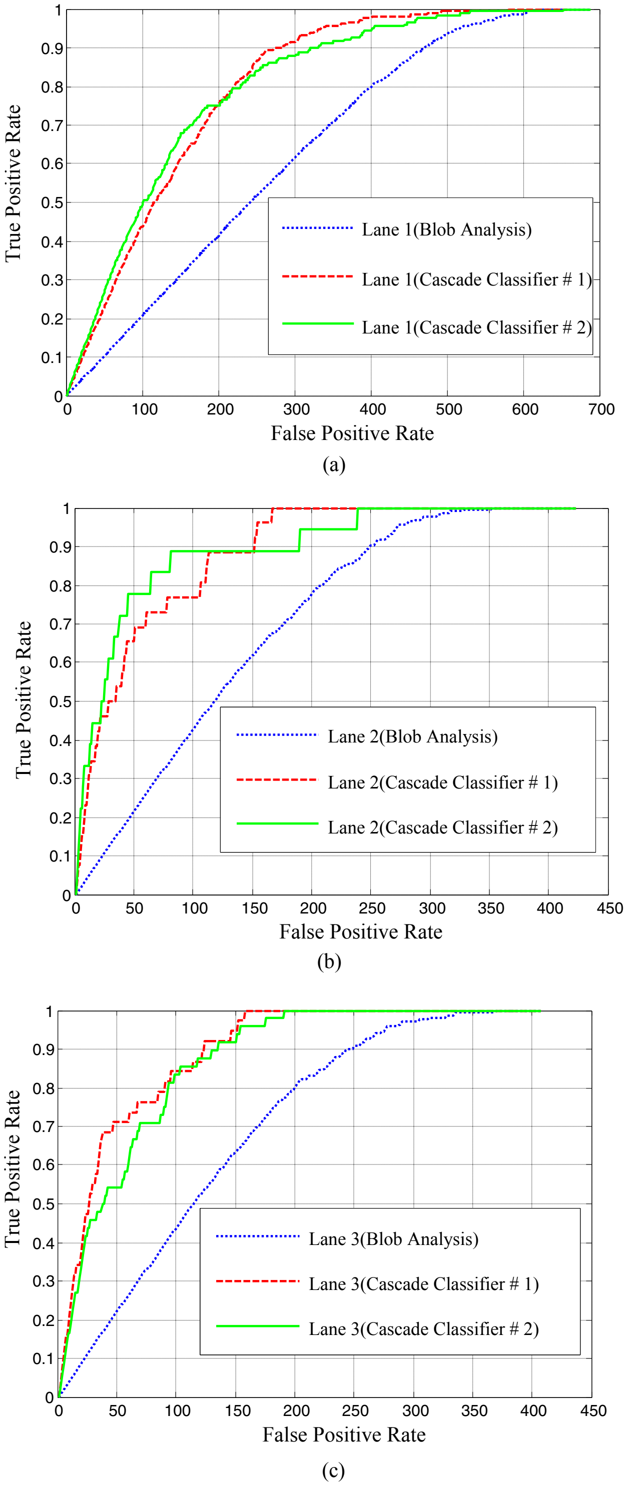


Fig. 4. Detection results obtained from Haar Cascade Classifier. Lanes are labeled as 1, 2 and 3 (a) NIPA (b) TOLL PLAZA Datasets

The detection performance of the used algorithms is analyzed by using Receiver Operating Characteristics curves (ROC). For NIPA dataset, the acquired results reveal that Haar cascade classifier has the highest accuracy as compared to blob analysis. The numbers of positive images and stages are kept higher in cascade classifier 2 in order to enhance detection results. ROCs of stated algorithms for lane 1, lane 2 and lane 3 are shown in Fig. 5. The detection results of each lane are analyzed individually. It has been noted that correct detection rate for lane 1 has comparatively low accuracy as compared to lane 2 and lane 3 for both methods. This is due being the farthest lane and having camera viewing angle which resulted in increased complexity and occlusion. Similarly, for TOLL PLAZA dataset, Haar cascade classifier performed better as compared to Blob analysis and reflected by ROCs in Fig. 6. Due to comparatively slow vehicle speeds and frontal view angle of the dataset thereby having better appearance and motion information both Haar classifier and blob analysis performed better in terms of detection rate as compared to NIPA results.

V. Conclusion

This paper provides an insight on the performance of motion (blob analysis) and appearance (Haar cascade classifier) based methods for vehicle detection in native environment. The dataset includes complex urban traffic scenes where road markings and signs are significantly low. In addition, type of vehicles ranges from motorbikes to long trucks having varying speeds thereby posing increased detection challenges. The detection results obtained from the trained Haar cascade classifier are more promising than blob analysis. Some false and misdetection are occurring due to varying vehicle speed, cast shadows, background clutter and camera viewing angle causing detection complexity and occlusion. As a future work, we plan to incorporate shadow modeling and elimination and enhance the feature pool that will help improve vehicle detection in the subject multifaceted environment.



(a)

(b)

(c)

Fig. 5. ROC Curve for NIPA Dataset (a) Lane 1 (b) Lane 2 (c) Lane 3

978-1-5090-4171-8/16/$31.00 ©2016 IEEE

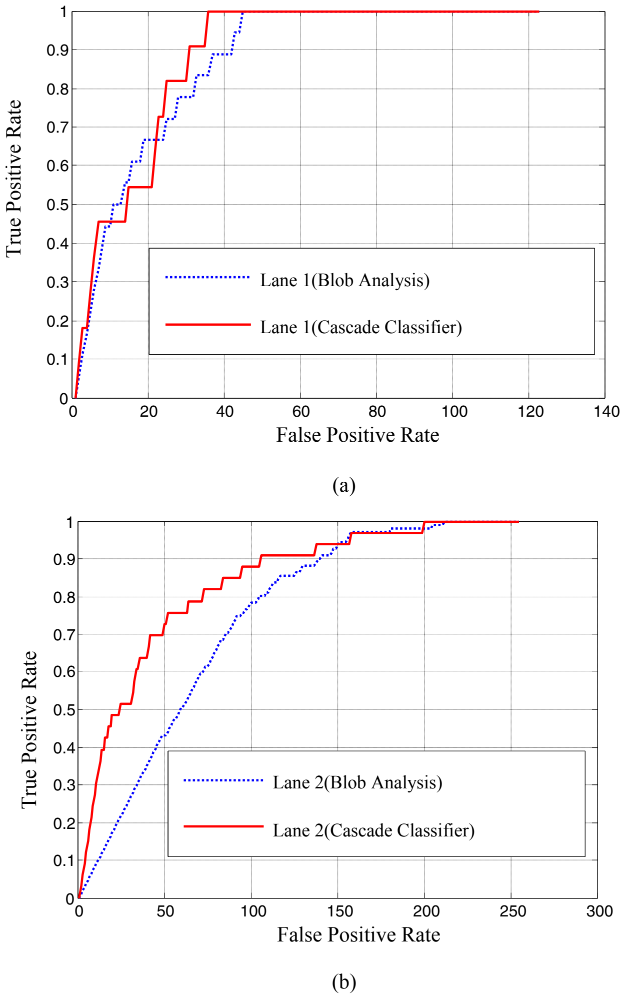


Fig. 6. ROC Curve for TOLL PLAZA Dataset (a) Lane 1 (b) Lane 2

References

[1] P. Viola, M. J. Jones, “Robust Real-time Object Detection,” International Journal of Computer Vision 57(2)”, 137–154, 2001.

[2] P. Menezes, J. C. Barreto, J. Dias, "Face Tracking based on Haar-like Features and Eigen faces." IFAC/EURON Symposium on Intelligent Autonomous Vehicles. Vol. 500, 2004.

[3] A. Broggi, A. Cappalunga, S. Cattani, P. Zani, “Lateral Vehicles Detection using Monocular High Resolution Cameras on Terra Max TM”, IEEE Intell. Veh. Symp., pp. 1143-1148, 2008.

[4] A. Vedaldi, B. Fulkerson, “VLFeat: An Open and Portable Library of Computer Vision Algorithms,” ACM international Conference on Multimedia, 2010.

[5] A. Jazayeri. H. Cai, J. Y. Zheng, M. Tuceryan, “Motion Based Vehicle Detection Identification in Car video”, IEEE Intell. Veh. Symp., 23, (3), pp. 493-499, 2010.

[6] R. Feris, J. Petterson, B. Siddiquie, L. Brown, S. Pankanti, “Large Scale Vehicle Detection in Challenging Urban Surveillance Environment”, Applications of Computer Vision (WACV) IEEE Workshop, 2011.

[7] S. Pankanti, L. Brown, J. Connell, A. Datta, Q. Fan, R. S. Feris, N. Haas, Y. Li, N. Ratha, H. Trinh, “Practical Computer Vision: Example Techniques and Challenges”, IBM J RES & DEV, vol 55 Paper 3, 2011.

[8] R. Feris B. Siddiqui, Y. Zhai, “|Attribute-based Vehicle Search in Crowded Surveillance Videos”, CMR’11 Proceedings of the 1st ACM International Conference on Multimedia Retrieval Article No. 18 ACM New York, NY, USA, 2011.

[9] N. C. Mithun, N. U. Rashid, S. M. Rahman, “Detection and Classification of Vehicles from Video Using Multiple Time-Spatial Images”, IEEE Transactions on Intelligent Transportation Systems, vol. 13, no. 3, 2012.

[10] J. P. Jodoin, G. A. Bilodeau, N. Saunier, “Urban Tracker: Multiple Object Tracking in Urban Mixed Traffic”, IEEE Winter Conference on Applications of Computer Vision, 2014.

[11] Y. Chen, G. Qin, “Video-Based Vehicle Detection and Classification in Challenging Scenarios”, International Journal on Smart Sensing and Intelligent Systems, Vol.7, No.3, 2014.

[12] M. U. Arif, Z. Lodhi, M. Khan, R. H. Raza, “Detection & Classification of Vehicles in Varying Complexity of Urban Traffic Scenes”, Electronic Imaging, Video Surveillance and Transportation Imaging Applications, pp. 1-9(9), 2016.

[13] X. Zhuang, W. Kang, Q. Wu, “Real-time vehicle detection with foreground-based cascade classifier”, IET Image Processing, vol 10, issue 4, 2016.

[14] https://github.com/sonots/tutorial-haartraining /tree/master/data/negatives [18-06-2016].

978-1-5090-4171-8/16/$31.00 ©2016 IEEE