An efficient vision-based traffic light detection and state recognition for autonomous vehicles

Conference Paper · June 2017

DOI: 10.1109/IVS.2017.7995785

Sanjay Saini
Harman International Industries Inc.

Krishna reddy Konda
ZF Friedrichshafen

Harish S Bharadwaj
Samvardhana Motherson Reflectec Automotive Systems India Limited, Bengaluru, In…

2017 IEEE Intelligent Vehicles Symposium (IV)
June 11-14, 2017, Redondo Beach, CA, USA

An Efficient Vision-Based Traffic Light Detection and State Recognition for Autonomous Vehicles

Sanjay Saini[1], Nikhil S[1], Krishna Reddy Konda[1], Harish S Bharadwaj[1] and Ganeshan N[1]

Abstract— **Traffic Light Detection(TLD) and understanding their state semantics at intersections plays a pivotal role in driver assistance systems and, by extension, autonomous vehicles. Despite of several reliable traffic light state detection approaches in literature, traffic light state recognition still remains an open problem due to outdoor perception challenge which includes occlusions, illumination and scale variations. This paper presents a vision-based traffic light structure detection and convolutional neural network (CNN) based state recognition method, which is robust under different illumination and weather conditions. In the first step, traffic light candidate regions are generated by performing HSV based color segmentation, which are then filtered out using shape and area analysis. Further, in order to incorporate the structural information of traffic light in diverse background scenarios, Maximally Stable Extremal Region (MSER) approach is employed, which helps to localize the correct traffic light structure in the image. To further validate the traffic light candidate regions, Histogram of Oriented Gradients (HOG) features are extracted for each region and traffic light structures are validated using Support Vector Machine (SVM). The state of the traffic lights are then recognized using CNN. To evaluate the performance of the proposed method, we present several results under a variety of lighting conditions in a real-world environment. Experimental result shows that the proposed method outperforms other vision based conventional methods under varying light and weather conditions.**

I. INTRODUCTION

Recently, research on intelligent vehicles, especially on autonomous vehicles in urban scenario has seen increased interest. Although, in near future fully autonomous driving is a possible scenario thus detection and interpretation of traffic light is a vital tool for autonomous vehicles. Many computer vision groups are working on this task in order to accomplish reliable traffic light detection. A wide variety of vision and active sensors (such as GPS, sonar, radar, and lidar) based approaches have been presented in literature [1], [2], [3] and –others continue to appear. However, detection and interpretation of traffic light for current advances in autonomous vehicle still remains an open problem, due to outdoor perception problems which consists of illumination variation effects and scale variations [1], [3].

The main aim of traffic light detection methods is to find the location of traffic light in each frame and understand their denotation. Most of vision-based methods have followed a common pipeline for traffic light recognition system: traffic light candidate region extraction, detection of traffic light structure and state recognition (type of signal). Recently, some sophisticated approaches have been presented in literature which employ external sensor information such as GPS, radar and lidar, to robustly find the potential presence of traffic light candidate regions in the image. Levinson et al. [4], presents a method for TL recognition by utilizing prior information about the location of traffic light and by utilizing probabilistic template matching in HSV color space to find the state of traffic light. Similarly, Firefield et al. [3], presents a method in which simple color segmentation based classifier has been used to find the colors of traffic light (i.e. red and green) within each predicted predefined bounding box. Although, information from external sensors improved the detection accuracy, the requirement of expensive hardware such as GPS or Lidar, leads to high-cost of the system. Further, feasibility of such systems is limited in current scenario, as it requires new infrastructure.

Recently, the color information of the light emitting units have been utilized to extract the traffic light candidate regions [5], [6], [7]. In order to improve the detection accuracy, shape information has also been incorporated with color information, including the well-known Hough Transform [7] and fast radial symmetry transform [8]. However, color based detectors rely heavily on color information, hence, leading to a high number of false candidates and below par performance in complex environments and adverse lighting conditions.

In this paper, we present a purely vision-based traffic light state detection method which is robust under different illumination conditions. The proposed method uses color segmentation and area based rejection filters for traffic light candidate region extraction. Localization of traffic light structure was done using MSER (Maximally Stable Extremal Regions) which has not yet been used for the task of traffic light detection. To further validate the potential traffic light candidate regions, HOG features were extracted and a non-linear SVM classifier was trained using the training examples. The detected traffic light structures were then used to determine the state of the signal using CNN based model.

The remainder of the paper is organized as follows: Section 2 gives a informative summary of the related work. Section 3 describes the proposed method for traffic light detection and recognition. Section 4 presents traffic light detection results and performance analysis of proposed method with existing potential vision-based methods. Section 5 contains conclusions and future scope for further research.

II. RELATED WORKS

During the last decade many researchers have made efforts to address the traffic light detection problem [1], [2], [3].

[1]Sanjay Saini, [1]Nikhil S, [1]Krishna Reddy Konda, [1]Harish Bharadwaj are with SMR Automotive Systems India Limited, Bangalore-560 100, India. sanjay.1987saini@gmail.com

However, most of these approaches are more or less defective in urban driving scenario. Most of the vision-based Traffic Light Recognition (TLR) systems presented in literature can be divided into four main stages: Traffic light(TL) detection, feature extraction, TL state classification and tracking. In this section, we present a brief summary of methods used in each of the stages.

As previously indicated, in the context of TLs, color is a major characteristic and it has been extensively used for preliminary reduction of search space in the image and for traffic light candidate region selection. Various types of color spaces have been discussed in the literature for traffic light detection. Several works have been presented in the literature based on normalized RGB color space like [7]. On the other hand, HSV color space has been widely used for traffic light detection[6], [5], as it is close to human vision and it is capable of efficiently detecting color distortion. Recently some of researchers have also explored other color spaces such as LAB [9], YCbCr [10] and LUV [1]. In order to make our system invariant to illumination and to reduce the number of feature spaces to 2 dimensions, we have used HSV based color segmentation as the initial stage of our algorithm.

Color segmentation generates a lot of false candidate regions. These regions can be eliminated based on various traffic light specific rejection filters such as aspect ratio, texture and size of detected blob. Generally, shape information can also be used after color segmentation for detection of traffic light. Hough transform was used on the edge map of a Laplacian edge detection filter [11]. A similar method was presented in [7], where circular Hough transform was used to find circles. Their results suggest that performance of circular Hough transform is better than conventional Hough transform. Fast radial symmetry has also been used to find the traffic light circles [8]. BLOB analysis based properties like size of detected blob, aspect ratio, shape of the blob etc., have been used by many researchers to eliminated the false candidate regions [5], [6].

Apart from color and shape information, structure information are also widely used as traffic light features. Mostly, the structure information of the traffic light are extracted using global and local image features descriptors. Many vision based feature descriptors have been successfully explored for traffic light detection including HOG [5], Haar-like-features [6], Hough Transform [7], 2D Gabor Wavelet feature [10] and Local Binary Pattern (LBP) [12]. However, traffic light detectors which rely heavily only on either of color, shape or structure information, leads to a below par performance in complex environments and adverse lighting conditions. The detection performance and robustness can be improved by using fusion of multiple features which comprise different information.

To recognize the state of the traffic light, different types of classifiers have been used in literature. However most of reviewed approaches use SVM, because of its ability to approximate highly non-linear problem, with relatively robust and reliable performance, and with relatively fewer hyper parameters. Chen et al.[5] uses HOG features to train SVM classifier to recognize traffic signals from the candidate regions. Similarly, Chiang et al.[12] used SVM to classify the state of TL from LBP features. AdaBoost cascade classifier has also been used for traffic light state classification using Haar features [6]. More recently, Convolution Neural Network (CNN) has been used to classify the traffic light states [9]. This method produces accurate results, but it is computationally very expensive. Several other classifiers have been utilized for the same task, the summary of this can be seen in [2]. As previously indicated, the classifier accuracy is heavily dependent on the quality of the image features which are extracted from candidate regions.

As can seen in [2], most of the presented approaches apply some form of tracking in order to improve the detection and classification of traffic lights. Tracking is generally important when, detection of traffic signal temporarily drops because the traffic signal may be occluded by external objects or under extreme lighting or weather condition. Different types of tracking has been used for traffic light detection and recognition including CAMSHIFT[6], Kalman Filter (KF) and Particle Filter(PF)[1].

III. Overview of the Proposed System

The main objective of this paper is to design an efficient purely vision-based traffic light detection and state recognition system. The pipeline of the proposed system is shown in Figure 1 and the steps are as follows.

- Candidate region extraction
  - Input images were pre-processed to limit the number of candidate regions using HSV based color segmentation method.
  - Further these candidates were filtered using aspect ratio and area based analysis.
  - MSER was applied on the candidate regions to localize the structure of the traffic light in diverse background scenarios.
- Detection of traffic light structure
  - HOG features were extracted from each candidate region.
  - Detection of traffic light structure was done based on extracted HOG features using SVM.
- Recognition of traffic light state
  - The state of the traffic light was recognized using CNN based methods.

A. Candidate Regions Extraction

Signalling systems in most countries are standardized to have a specific type of traffic light structure and color. Traffic lights are characterized by specific colors: Red, Amber and Green. Color is a major characteristic in traffic light, thus color segmentation is a very important step in our proposed system for limiting the number of potential traffic light based candidate regions. In RGB color space, the luminance and chrominance components are not decoupled, hence it is not illumination invariant. Traffic lights are exposed to varied lighting conditions, therefore it is very important to make

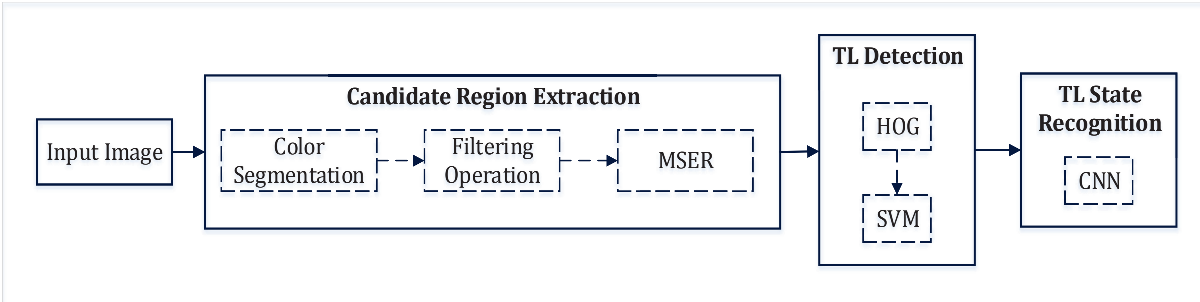
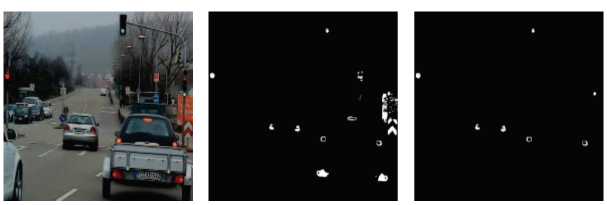


Fig. 1: Overview of the proposed system

the system invariant to illumination. Hence we used HSV color space which separates luminance and chrominance components. Hue and saturation are the two features that we are more interested in HSV color space, thereby also reducing the feature space from 3-D in RGB color space to 2-D in HSV space.

Color segmentation technique results in detection of objects which have chrominance similar to that of traffic light, thereby yielding many false positives (non-traffic attributes). Majority of these false positives are eliminated by applying traffic light specific rejection criteria (i.e. aspect ratio and area of the detected candidates). At the end of this process substantial amount of false positives were eliminated, however many false positives still remain and they will be removed using shape based features as explained in next section. The outcome of color segmentation and area based filtering process is shown in Figure 2.

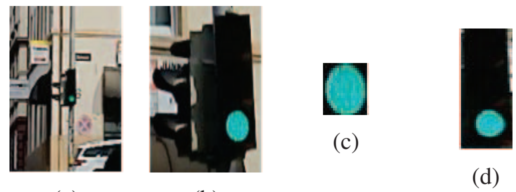


(a) Input Image (b) Segmentation (c) Filtering Process

Fig. 2: Color segmentation and filtering process

The contour detected using color segmentation contains traffic light and other similar colored objects. Size of the traffic light is pretty small and number of features resulting from traffic light is relatively insufficient to represent a complex structure of the traffic light. Therefore, we extract the entire traffic light structure for more relevant features, which results in more efficient detection and recognition phases. Figure 3(c) shows a contour containing a traffic light, whereas Figure 3(d) shows a contour containing an entire traffic light structure. Extraction of traffic light structure is done by initially fitting a bounding box over the traffic light. The size of the bounding box was fixed based on the maximum size of the traffic light structure that can be encountered by the system. The maximum size of the bounding box was determined from ground-truth.

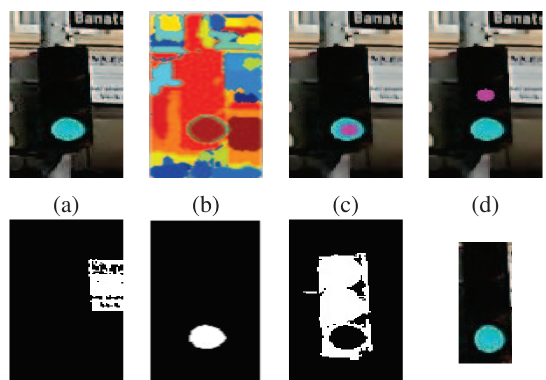
Figure 3(a) and 3(b) depict the area covered by a fixed size bounding box when the signal is far from camera and when the signal is close to camera. In absence of a localization method for traffic light structure, the bounding box for a



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

Fig. 3: Example of fixed size bounding box

signal which is far from camera will contain traffic signal along with diverse background which are non-traffic regions. This makes it difficult to effectively detect and recognize the state of a traffic signal. Thus it is very important to localize the traffic light structure effectively.



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

(e) (f) (g) (h)

Fig. 4: Traffic light localization process using MSER

To localize the correct structure of a traffic light in diverse background and to make the size of the bounding box adaptable based on the distance of camera from the signal, Maximally Stable Extremal Regions (MSER)[13] technique has been applied. Figure 4(b) shows the result of MSER clustering. In MSER, each frame is binarised at a number of different threshold levels. The regions which remain stable over large range of thresholds are the regions of interest. Traffic light structures undergo a lot of affine transformation with respect to camera and the size of traffic lights vary considerably with respect to viewpoint. MSER regions are invariant to affine transformation as well as they support multi-scale detection (capable of detecting both fine and large structures). Both these properties of MSER are important characteristics for traffic light structure localization.

MSER results in various clusters in the image as shown in Figure 4(e-g). To select the most relevant clusters, we select two seed points. One seed point corresponding to the centre of the signal, as detected using color segmentation and another seed point corresponding to any point of the signal structure. Figure 4(a), shows the input image and Figures 4(c) and (d) indicate the seed points 1 and 2 respectively. The clusters that do not correspond to the seed points are eliminated (example, cluster in Figure 4(e)). Next step is to shortlist the clusters which best represents a traffic signal. Clusters which correspond to seed point 1 (Figure 4(f)) and seed point 2 Figure 4(g) are selected, then the two clusters are integrated and the resultant output is a contour corresponding to the localized traffic signal structure as shown in Figure 4(h). The outcome of this stage is processed in next stage for detection and recognition of traffic light.

B. Traffic Light Structure Detection

For each of the candidate regions resulting from the previous step, Histograms of Oriented Gradients (HOG) features were extracted to aid in the detection of the traffic light structures, thereby eliminating most of the false positives. HOG was introduced by Dalal et al. in [14]. The feature was developed with the aim of detecting humans in an image. Later on, the utility of the feature was extended to pedestrian detection, object detection and other computer vision problems. HOG features are relatively invariant to scale and rotation which is important for traffic lights. HOG features are computed by taking orientation histograms of edge intensity in a local region. As indicated previously, color is a major characteristic of a traffic signal. Many researcher suggested [15] that performance of the conventional HOG can be improved by combining HOG features over multiple color channels. Therefore, in this paper we have incorporated color information with HOG feature as described in [15]. A simple visualization of HOG feature extraction is illustrated in Figure 5. For extracting the HOG features, the detection window size was fixed at $(width \times height = 90 \times 180)$ empirically.

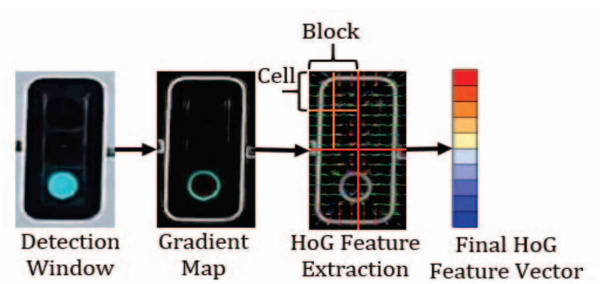


Fig. 5: HoG Feature Extraction

HOG feature descriptors are then fed to a non-linear SVM classifier for detecting traffic light structures. Non-linear SVMs will create a space transformation, it will be a linear SVM in the feature space, but a non-linear separation border in the input space. Lower the number of input features, easier it is for the non-linear SVM to perform space transformation. Due to large amount of training data and relatively small number of HOG features, we chose non-linear SVM classifier. SVM is a supervised learning model, it constructs a hyper-plane in a high dimensional space to separate the feature points into two or more classes. The feature points from which the separated hyperplane is located at the maximum margin are known as support vectors. For a test data consisting of a potential candidate regions, the HOG features are extracted and classified by calculating the distance between the extracted feature points of the test image with the support vectors found during training phase.

C. Traffic light state recognition

In recent years deep neural networks and in particular convolutional neural networks(CNN) have been very successful in solving computer vision problems as they rely on very basic features to obtain high level abstraction of training data. Thereby achieving high accuracy in classification. Hence we have opted for CNN based classifier for traffic light state recognition. Proposed network architecture is derived from CNN introduced by Yann LeCun [16] for handwritten digit recognition. However by incorporating application specific modifications in the network, we have been able to reduce the number of parameters to 76000 from the original 1400000 which amounts to reduction of complexity by about 95 percent. Block diagram of the network is given in Figure 6. Output of detection algorithm is resized to 48×96 pixel size before being given as an input to CNN. Initial convolution layer consists of 20 filters of size 5×5. Filter stride across the input image during each iteration is 2 pixels. After convolution layer we have a max pool layer with a pixel stride of 2. Max pool layer is followed by a second convolution layer with 50 filters of 5×5 with a stride length of 2 pixels. Output convolution layer is followed by second pooling layer. After the second pooling layer two consecutive fully connected layers map the network to three neurons which correspond to three classes of traffic lights. Parameter reduction is primarily achieved by reducing number of filters and increasing the stride length to two. Since traffic lights consist of limited number of shape related features, accuracy is not affected by such modifications as can be seen from Table II.

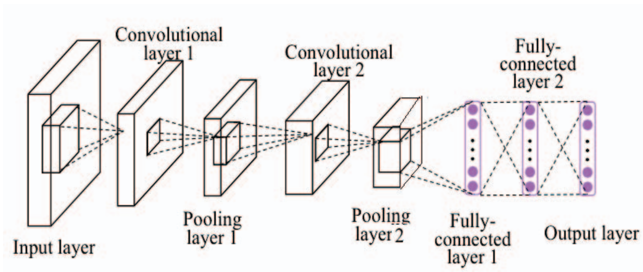


Fig. 6: CNN architecture for TLs state recognition

IV. Experiment Results

Quantitative evaluation has been performed on a common set of video sequences to study the performance of the proposed traffic light detection and state recognition algorithm as against other state of the art approaches. The utilized video dataset was captured in Germany under varying lighting and weather conditions using a planar camera with a field of view of around 40 degrees and the camera was placed on the dashboard of the recording car. The video set consists of 36000 frames which has 8154 instances of different traffic lights (that are red, green and amber), as marked from ground truth. Among them, 4000 samples of traffic light were randomly selected as positive training data for TLD and 12000 samples were selected as negative samples (non-traffic light instances). The variability in the pattern/structure of non-traffic lights is comparatively higher than that of traffic lights. Hence the number of non-traffic light contours used as training data for TLD were relatively higher. For the task of traffic light state recognition, CNN was trained with 4099 and 2443 green and red signal samples respectively. The performance of different TLD algorithms were evaluated using the following steps:

- For the training and testing a common set of image sequence is selected.
- The detection output is quantified with the manually annotated ground truth.
- Each of the candidates are placed under one of the 4 classes - True Positive(TP), False Positive(FP), True Negative (TN) and False Negative (FN), based on the comparison with ground truth.
- A set of statistics (Precision, Recall and F-measure) is calculated for each implemented algorithm.

Each of the TL detection algorithm is evaluated based upon the following information retrieval measurements:

- Precision and Recall, as seen in equation 1
- F-measure, as seen in equation 2

Precision and Recall quantify how well the detected TLs matches with the ground-truth. High precision or high recall value means high performance. F-measure, indicates the overall accuracy of the system.

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN} \tag{1}$$

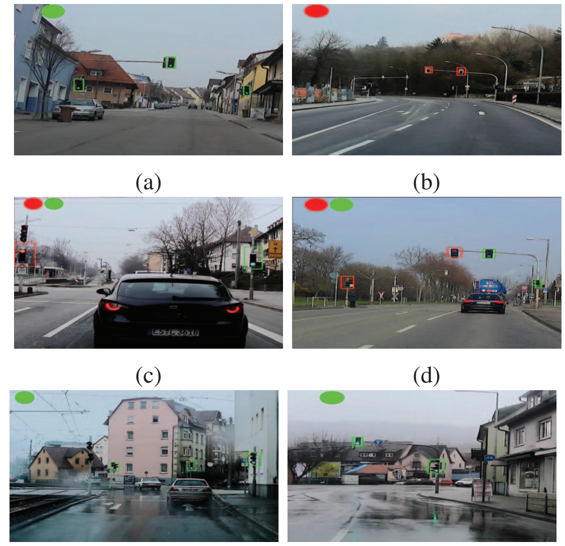
Precision is the ratio of correct TL detections compared to the total number of detections. True positives indicate the number of traffic signals which were rightly classified as a traffic signal. False positives indicate the number of non-traffic signals classified as a traffic signals. Recall is the ratio of correct TL detections compared to the actual number of TLs.

$$F = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{2}$$

F-measure is the geometric mean of precision and recall.

A. *Quantitative and visual results*

The proposed algorithm was tested under varying lights and different weather conditions. We noticed that proposed method was able to work consistently under different illumination and weather conditions. Figure 7 demonstrates the visual outcome of the proposed method for traffic light recognition in different environment conditions. The bounding box around the traffic light indicates the detection results and solid circle in the left corner of the image shows the state recognition results. The detection and recognition of the multiple traffic lights in urban scenario is shown in



(a) (b) (c) (d) (e) (f)

Fig. 7: Visualization of automatic traffic light detection in various scenarios

Figure 7 (a)-(d) while Figure 7(e)-(f) shows the detection and recognition of traffic lights in rainy conditions.

Table I and II, shows the traffic light detection and recognition accuracy in greater details indicating the difference in values between the ground-truth and the automatic detection in each frame.

TABLE I: Performance of the proposed method with existing state of the art methods (4154 test samples)

Methods	Precision	Recall	F-measure
SIFT+BoF+SVM	93.13	85.86	89.35
HOG+Linear SVM	94.14	94.39	94.26
HOG+ Non-Linear SVM	98.81	92.24	95.41
MSER+HOG+Non Linear SVM	99.41	98.65	99.03

TABLE II: Performance of CNN and SVM for TL state recognition

	Green			Red		
Methods	Total	Correct	Efficiency	Total	Correct	Efficiency
CNN	4099	4090	99.78	2443	2427	99.35
SVM	4099	4006	97.73	2443	2308	94.47

An efficient TL detector should rely on gradient magnitude and orientation as well as possess relative invariance to scale, 3-D viewpoint and illumination. In this paper, for the task of detecting traffic light structure, we made a comparison between two popular feature descriptors, Scale invariant feature transform (SIFT) and HOG. SIFT is a local descriptor of image features and does not provide an effective representation of entire structure of traffic light. Using a set of such local descriptors rather than a single descriptor would provide a more global representation of a traffic light. Hence, Bag of Features (BoF) was used to generate an aggregate representation of the occurrence of multiple features in an image.

Experimental results shows that HOG features outperform SIFT features for traffic light structure detection (using linear SVM). The main reason being, HOG as a global descriptor encodes gradient information of the region in and around the traffic light, thus it is able to describe the traffic light structure more effectively. HOG features are extracted using sliding window, whereas SIFT features are extracted by intially detecting keypoints/feature points and then describing the detected feature points. For traffic lights which are relatively small sized contours, HOG is computationally faster than SIFT. HOG takes around 0.5ms to extract features from a single contour, whereas SIFT takes around 2ms per contour.

We also compared the performance of linear and nonlinear SVM classifier for TL structure detection. Results show that though the number of support vectors were relatively high in case of non-linear SVM, the accuracy of non-linear SVM was comparatively better than that of linear SVM. Owing to the size of the traffic signal, the dimension of the HOG descriptor is relatively small. In the presence of higher number of samples for training (as in our case), with relatively fewer features, non-linear SVM was more effective for the task of detecting traffic light structures.

Traffic lights are extremely transformed and the size of the traffic signal not just depends on the distance from camera, but varies considerably with respect to 3-D viewpoint and distortion caused due to motion or due to lighting conditions. On the hindsight, signal structure might be comparatively less variant to 3-D viewpoint changes. Hence, performing localization of traffic signal structure based on the size of the traffic light detected using color segmentation was not very robust. It was noted through experimental results that the performance of the system improved considerably by using MSER for localizing the traffic signal structure. Through experimental analysis as shown in Table I, it was ascertained that using MSER, HOG features and non-linear SVM, the performance of the system for detecting traffic signal structures was found to be robust under varied lighting conditions, weather condition and 3-D viewpoint achieving an F-measure score of 99.03%.

As can be seen in Table II, CNN performance is better than SVM for the task of traffic light recognition. This is because, CNN consists of several convolutional layers and fully connected layers, that learn a better representation of the input data, concentrating abstract features rather than class specific features.

In terms of computational complexity, the run time of the entire system was approximately 39 frames per second on a 3.6 GHz Xeon Processor aligned with NVIDIA Quadro K2200 GPU for an image resolution of 1280 x 800. Candidate region extraction and TL detection block as shown in Figure 1 are run on PC platform and take around 8 and 15 milliseconds respectively. CNN for TL state recognition runs at around 3 milliseconds per frame on GPU using Caffe framework.

V. CONCLUSIONS

This paper presents a purely vision-based real-time traffic light detection and state recognition method which is robust under different illumination conditions. In the proposed method, initial traffic light candidate regions are selected using color segmentation. We present a novel traffic light structure localization from image by using MSER. In order to further validate the traffic light and to reduce the false positives, HOG features and non-linear SVM have been used. State of the traffic light has been recognized using CNN based classifier. Experimental results suggest that the proposed method is consistent under different illumination and weather conditions, without specific tunning in the parameters. Results also shows that accuracy and consistency of the proposed method are better than conventional approaches.

REFERENCES

[1] M. Diaz, P. Cerri, G. Pirlo, M. A. Ferrer, and D. Impedovo, "A survey on traffic light detection," in *International Conference on Image Analysis and Processing*. Springer, 2015, pp. 201–208.

[2] M. B. Jensen, M. P. Philipsen, A. Møgelmose, T. B. Moeslund, and M. M. Trivedi, "Vision for looking at traffic lights: Issues, survey, and perspectives," 2015.

[3] N. Fairfield and C. Urmson, "Traffic light mapping and detection," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 5421–5426.

[4] J. Levinson, J. Askeland, J. Dolson, and S. Thrun, "Traffic light mapping, localization, and state detection for autonomous vehicles," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 5784–5791.

[5] Q. Chen, Z. Shi, and Z. Zou, "Robust and real-time traffic light recognition based on hierarchical vision architecture," in *Image and Signal Processing (CISP), 2014 7th International Congress on*. IEEE, 2014, pp. 114–119.

[6] J. Gong, Y. Jiang, G. Xiong, C. Guan, G. Tao, and H. Chen, "The recognition and tracking of traffic lights based on color segmentation and camshift for intelligent vehicles." in *Intelligent Vehicles Symposium*, 2010, pp. 431–435.

[7] M. Omachi and S. Omachi, "Traffic light detection with color and edge information," in *Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on*. IEEE, 2009, pp. 284–287.

[8] S. Sooksatra and T. Kondo, "Red traffic light detection using fast radial symmetry transform," in *Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2014 11th International Conference on*. IEEE, 2014, pp. 1–6.

[9] V. John, K. Yoneda, B. Qi, Z. Liu, and S. Mita, "Traffic light recognition in varying illumination using deep learning and saliency map," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2014, pp. 2286–2291.

[10] Z. Cai, Y. Li, and M. Gu, "Real-time recognition system of traffic light in urban environment," in *2012 IEEE Symposium on Computational Intelligence for Security and Defence Applications*. IEEE, 2012, pp. 1–6.

[11] E. Koukoumidis, M. Martonosi, and L.-S. Peh, "Leveraging smartphone cameras for collaborative road advisories," *IEEE Transactions on mobile computing*, vol. 11, no. 5, pp. 707–723, 2012.

[12] C.-C. Chiang, M.-C. Ho, H.-S. Liao, A. Pratama, and W.-C. Syu, "Detecting and recognizing traffic lights by genetic approximate ellipse detection and spatial texture layouts," *International Journal of Innovative Computing, Information and Control*, vol. 7, no. 12, pp. 6919–6934, 2011.

[13] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and vision computing*, vol. 22, no. 10, pp. 761–767, 2004.

[14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 886–893.

[15] K. Van De Sande, T. Gevers, and C. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 9, pp. 1582–1596, 2010.

[16] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, "Handwritten digit recognition with a back-propagation network, 1989," in *Neural Information Processing Systems (NIPS)*.