Ongoing Work on Traffic Lights: Detection and Evaluation

Mark P. Philipsen
Aalborg University
Denmark
markpp@gmail.com

Morten B. Jensen
Aalborg University
Denmark
mbornoe@gmail.com

Mohan M. Trivedi
UC San Diego
USA
mtrivedi@eng.ucsd.edu

Andreas Møgelmose
Aalborg University
Denmark
am@create.aau.dk

Thomas B. Moeslund
Aalborg University
Denmark
tbm@create.aau.dk

Abstract

Research in traffic light recognition (TLR) has stagnated compared to related computer vision areas, such as pedestrian detection and and traffic sign recognition. We focus on the detection sub-problem, since this is the most challenging problem and solving this is the key to a successful TLR system. This is done by looking at four detectors from different author groups and their reported results. From surveying existing work it is clear that currently evaluation is limited primarily to small local datasets. In order to provide a common basis for future comparison of TLR research an extensive public database is collected based on footage from US roads. The database consists of continuous test and training video sequences, totaling 46,418 frames and 112,971 annotated traffic lights. The sequences are captured by a stereo camera mounted on the roof of a vehicle driving under both night and day conditions with varying light and weather.

1. Introduction

Driver assistance systems are gaining a lot of momentum currently, as evident in top models from prominent car manufacturers. Recognition of traffic lights (TLs) would be a desirable addition, but judging by the state of current research in this area, consumer-ready systems are not on the immediate horizon. Traffic light recognition (TLR) consists of three sub-problems, detection, classification, and tracking. Figure 1 illustrates the typical flow of such a computer vision system. A similar breakdown is done for traffic sign recognition in [13].

The detection and classification stages are executed sequentially on each frame, whereas the tracking stage feeds back spatial and temporal information between frames. For TLR both the detection and classification stages are comparable to the equivalent stages in traffic sign recognition. Tracking of TLs differs, since signs are static and TLs change states. More about the coventions, structure and dynamics of TLs in section 2. The detection problem covers locating desired candidate TLs. Candidates are either rejected or accepted in the classification stage based on features extracted from the detected candidates. Furthermore, the state of accepted candidates is determined. In tracking the location and state of TLs are tracked through a frame sequence. Since detection of TL candidates is the foundation for a successful classification and tracking we will focus exclusively on this for the remainder of this paper. Unlike for sign recognition and pedestrian detection, no surveys of TLR research exist.

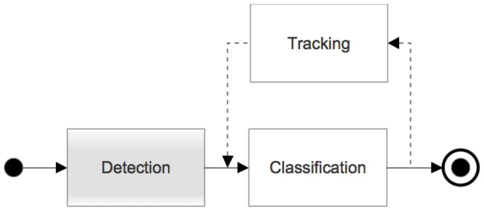


Figure 1: Breakdown of a vision based TLR system.

The purpose of this paper is to highlight some prominent approaches to TL detection from a few recent papers, as well as describe a common procedure for evaluation of such detectors. Most current research is evaluated based on local datasets with a limited number of TLs and little variation. This makes comparison between existing methods and new contributions difficult. We introduce a comprehensive TL dataset along with a proposal for a common evaluation procedure for TL detectors. The KITTI Vision Benchmark Suite [10] is an example of a dataset and benchmark used for evaluating various vision applications, such as stereo, object detection, and visual odometry, with the purpose of reducing the bias and providing real-world test scenarios. The purpose is similar for the recently introduced *VIVA Challenge*, which is used for benchmarking proposed methods on difficult tasks associated with drivers, occupants, vehicle dynamics, and vehicle surroundings. For now, it includes datasets for hands, faces, and signs captured under challenging naturalistic drive settings. The TL dataset published with this work is eventually going to be included in the *VIVA Challenge*[12].

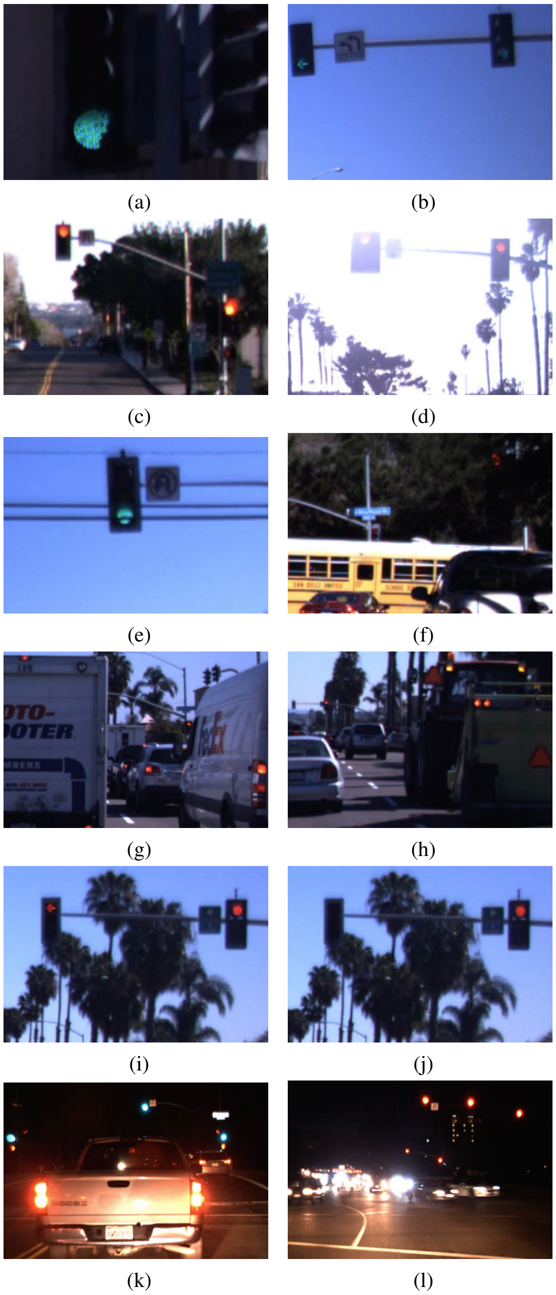
The contributions made in this survey paper are thus threefold:

1. Provide an overview of four different approaches to TL detection from four author groups.
2. Introduce a common evaluation procedure for TL detectors.
3. Publish an extensive high resolution, stereo video database, with day and night video sequences, accompanied by TL annotations.

The paper is organized as follows: Section 2, explains the possible appearances of TLs, along with common challenges that TL detection systems are subject to. Related research is summarized in section 3. In section 4, we present a new database for evaluation of TL detection systems. Finally, concluding remarks are found in section 6.

2. Traffic Lights: Structure and Challenges

TLs are by design made to stand out and be easily visible. Their primary components are bright colored lamps. These lamps are commonly circular or arrow shaped, they are surrounded by a uniform, often dark box. The purpose of TLs is the same across the world, they must safely regulate the traffic flow, by informing drivers about the the right of way. Right of way is given in a manner which minimize conflicts between vehicles and pedestrians traveling incompatible paths through the intersection during the same time span. The most common TL configuration is the basic red-yellow-green signal, where each state indicates whether a driver should stop, be prepared to stop, or keep driving. Worldwide there are many variations in TL designs; however, all follow a few general guidelines. A TL consists of a box that holds differently colored, and sometimes differently shaped lamps. The orientation, color, size, and shape of the box will vary country to country and even city to city. In the U.S. TLs are regulated by the Federal Highway Administration in the *Manual on Uniform Traffic Control Devices* [8] and most European countries have signed the *Vienna Convention on Road Signs and Signals* [18], requiring TLs to meet a common international standard.



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l)

Figure 2: (a) Examples of TLs from the collected dataset.

2.1. Challenges in recognizing traffic lights

Although TLs are made to be easily recognizable, influences from the environment and sometimes sub-optimal placement can make successful detection difficult, if not impossible. Issues include:

- Color tone shifting and halo disturbances because of influences from the atmosphere and glass that the light passes through[3]. Fig. 2(c).
- Occlusion and partial occlusion because of other ob-

jects or oblique viewing angles[3]. This is especially a problem with supported TLs [9, 5, 2]. Fig. 2(e),(f),(g).

- Incomplete shapes because of malfunctioning lights[3] or dirty lamps 2. Fig. 2(a),(b).
- False positives from, brake lights, reflections, billboards[7, 11], and pedestrian crossing lamps. Fig. 2(h).
- Changes in lighting due to adverse weather conditions and the positioning of the sun and other light sources. Fig. 2(d),(k),(l).
- Mismatch between camera's shutter speed and TL LED's duty cycle. Fig. 2(i),(j).

Inconsistencies in TL lamps can be caused by dirt, defects, or the relatively slow duty cycle of the LEDs. The duty cycle is high enough for the human eye not to notice that the lights are actually blinking. Issues arise when a camera uses fast shutter speeds, leading to some frames not contain a lit TL lamp. Saturation is another aspect that can influence the appearance of the lights. When transitioning between day and night, the camera parameters must be adjusted to let the optimal amount of light in and avoid under or oversaturation. [6] introduces an adaptive camera setting system, that change the shutter and gain settings based upon the luminosity of the pixels in the upper part of the frame.

3. Related Work

The four approaches which are examined and compared are selected based on being recent and by representing a niche in current TL detection.

The first approach of interest is found in [17, 2, 5] which share the same TL detector. The detector is first presented in [5] and it relies purely on intensity from grayscale images. This has the advantage of being more robust to color distortion. Areas brighter than their surroundings are segmented using the white top-hat morphology operation, which leads to an initial high number of false candidates. False candidates are filtered out based on the shape information. Specifically, rejection is done based on criteria such as, dimensional ratio, the BLOB being hole free and approximately convex. Furthermore, the areas of BLOBs are compared to the areas of regions grown from extrema in the original grayscale image. This is especially effective for removing false candidates found in bright uniform areas such as the sky. This detector relies heavily on a competent classifier for further rejection and state estimation, since the number of false candidates is very high and color information is not available. The detector reaches a detection rate of more than 90%. The detection rate is described as the ratio between correct TL recognitions and the ground truth in a given video sequence

[19] begins by detecting the vanishing line and thereby reducing the search area considerably, relying on the assumption that TLs will only appear above this line. They then apply the white top-hat operation, similarly to [17, 2, 5]. This is done on the intensity channel V from the HSV color space. The top-hat operation yields a greyscale likelihood map which is thresholded to create a binary candidate map. Remaining candidates are filtered based on statistical measurements of the hue and saturation ranges of red and green lights. All pixels outside these ranges are rejected while the remaining pixels are selected as candidates. Candidate BLOBs assigned scores based on size and height-width ratio. They then look for black bounding boxes around the BLOBs based on gradient information and the blackness of the inside of box candidates. These metrics are also translated into a score and combined with the BLOB score. Finally, candidates are confirmed if their scores are high. Their system reaches an accuracy of 85%. The accuracy is the overall recognition rate taking true positives, false negatives, and false positives into consideration.

[11] extracts candidate BLOBs from RGB images by applying a color distance transform proposed in [15]. The transform emphasizes the chosen color in an intensity image, which is thresholded to remove to suppressed colors. This is followed by shape filtering to reduce noise using width/height ratio and the solidity of BLOBs. The solidity is calculated based on the ratio between the area of the BLOB and it's bounding box. When evaluating their system, they count a success if the TL was detected just once in the sequence, this allows them to reach a detection rate of 93%.

An easy way of improving the segmentation is to reduce the search area. A popular and simplistic approach is to limit the search to the upper half of the input image. A sophisticated and precise approach is seen in [7], where an off-line database containing prior knowledge of TL locations is used. The off-line database is created using the input image combined with accurate GPS measurements, and then manually hand-labeling of the areas with TLs on a precaptured image sequences. Given that the route which the TLR system is used upon is hand-labeled, such an off-line database can reduce the searching window and the detection problem significantly. However, collecting and maintaining such a database is very time-consuming. Besides the use of prior knowledge, [7] uses color and shape information for final localization of candidates. With the use of their prior map their approach reach a precision of 99% and a recall of 62%.

Table 1 provides an overview of the four mentioned TL detectors along their reported performance measurements. Note that detector performance is reported using different measures and on vastly different data sets, making comparison difficult.

Table 1: Overview of recent studies in TLR. GT is an abbreviation of ground truth

Paper	**Year**	**Color Space(s)**	**Segmentation**	**Filtering**	**Dataset(GT TLs)**	**Conditions**	**Results**
[17, 2, 5]	2009 2009 2014	Grayscale	Top-hat spot light detection	BLOB filter, shape, structure	10,339 (*ParisTech*)	Urban day	90% Detection rate
[19]	2014	HSV	Top-hat spot light detection	BLOB filter, color, shape, structure	3,767	Good, challenging, very challenging	85% Accuracy
[11]	2013	RGB	Color thresholding	BLOB filter, shape	12,703	Night	93% Detection rate
[7]	2011	-	Prior knowledge of TL location	BLOB filter, color and shape	1,383	Morning, afternoon, night	99% Precision, 62% Recall

4. LISA Traffic Light Database

As it was concluded in the traffic sign survey paper [13], the general approach for testing and validating a proposed method is to use a privately collected dataset. This is considered sufficient for preliminary testing and validation. But, when trying to estimate the performance of a contribution, it becomes difficult to compare the work with others'. The only currently public dataset is provided by *Robotics Centre of Mines ParisTech* in France. The dataset consists of 11,179 frames from a single 8m 49s long video. It contains 9,168 hand-labeled instances of TLs. More information about this dataset can be found in Table 2.

A public database for computer vision should support the three part stereo vision bottom-up paradigm described in [16]. [16] provides an overview of vehicle detection systems based on both monocular and stereo vision since 2005. Both monocular and stereo vision are widely used for solving this problem, but an interesting finding in this work is the stereo vision bottom-up paradigm which consist of visual odometry, feature points in 3D, and distinguishing static from moving points, which is also mentioned in [4]. The motivation for having a public TL database with stereo images is therefore that all three parts in this paradigm can help reduce the amount of false positives. This notion is reinforced in [1] where the main technical challenges in urban environments are occlusions, shadow silhouettes, and dense traffic. The introduction of stereo has shown promising result in relation to solving these challenges.

The database that is collected and released together with this paper is focused on TLR and contains TLs that are found in San Diego, USA. Even though it was collected with TLR in mind, it could also become useful for evaluation of other related computer vision challenges, since it contains numerous traffic signs, vehicles, pedestrians, etc. The stereo image pairs are acquired using the Point Grey's three lens CCD camera, Bumblebee XB3 (BBX3-13S2C-60) with a resolution of 1280 x 960, each lens has a horizontal Field of View(FoV) of 43°and a focal length of 6mm. The stereo camera supports two different baselines, 12 and 24 cm, whereof a baseline of 24 cm is used for the LISA TL database. The stereo images are uncompressed and rectified on the fly, and captured with a frame rate of 16 FPS. Capturing was done by mounting the stereo camera centrally up front on the capturing vehicle's roof. The database provides two day and two nighttime sequences for testing and 18 shorter video sequences intended for training and additional testing. They are organized as seen in Table 3, which gives a detailed overview of all the video sequences that are made available with this paper. The number of annotations is the accumulated number of hand-labeled TLs on a frame-by-frame bases. The number of TLs is the physical number of TLs in the physical world. Camera gain and shutter speed were manually set to avoid oversaturation as well as to minimize flickering from the TLs. For all day clips, a shutter speed of 1/5000s and 0 gain was used. For all night clips, a shutter speed of 1/16s and 8 gain was used. A Triclops calibration file is provided along with the stereo images. This file contains the factory calibration for the used Bumblebee XB3 camera, which can be used with Point Grey's Triclops SDK.

Each sequence in the database comes with 2 hand labeled annotations for the left stereo frame. The annotations for a given video sequence contains the following information: frame number, rectangular area around the lit TL lamp or TL box, and the state of that area. Labeling is done for every visibly lit TL lamp. An example of annotated TLs is seen in Figure 3. The purple annotations are available which only mark the lit TL lamp. The green annotations cover most of the TL box and are meant for detectors that aim at detecting the entire box. The LISA Traffic Light Database is made freely available at `http:`

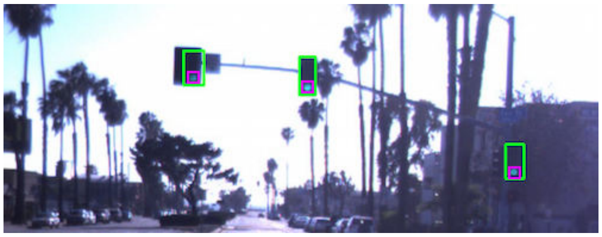


Figure 3: Example of annotated TLs.

Table 2: Overview of current public TL databases. Ambiguous means that it could not be decided whether the light was a TL during annotation, these are ignored when evaluating.

	***Robotics Centre of Mines ParisTech*[14]**	***LISA (Laboratory for Intelligent and Safe Automobiles) Traffic Light Database***
#Classes	4 (green, orange, red & ambiguous)	7 (go, go forward, go left, warning, warning left, stop, & stop left)
#Frames / #Annotations	11,179 / 9,168	46,418 / 112,971
Image spec.	640 x 480, 8-bit, RGB	Stereo, 1280 x 960, 8-bit, RGB
Place of origin	Paris, France	San Diego, USA
Video included	Yes, 8min 49s @25FPS	Yes, 44min 24s @16FPS
Description	1 urban day time sequence	4 test sequences $\geq$ 5min and 18 clips $\leq$ 2min 49s, morning, evening, night

Table 3: Overview of the video sequences in LISA Traffic Light Database.

Sequence name	**Description**	**# Frames**	**# Annotations**	**# TLs**	**Length**	**Classes**
Day seq. 1	morning, urban, backlight	4,800	10,267	25	5min	Go, warning, warning left, stop, stop left
Day seq. 2	evening, urban	9,586	11,154	29	6min 10s	Go, go forward, go left, warning, stop, stop left
Night seq. 1	night, urban	4,992	18,889	25	5min 11s	Go, go left, warning, stop, stop left
Night seq. 2	night, urban	6,533	23,776	54	6min 48s	Go, go left, warning, stop, stop left
Day clip 1	evening, urban, lens flare	2,161	6,474	10	2min 15s	Go, stop
Day clip 2	evening, urban	1,031	2,230	6	1min 4s	Go, go left, warning left, stop, stop left
Day clip 3	evening, urban	643	1,087	3	40s	Go, warning, stop
Day clip 4	evening, urban	397	859	8	24s	Go
Day clip 5	morning, urban	2,667	9,717	8	2min 46s	Go, go left, warning, warning left, stop, stop left
Day clip 6	morning, urban	468	1,215	4	29s	Go, stop, stop left
Day clip 7	morning, urban	2,718	8,189	10	2min 49s	Go, go left, warning, warning left, stop, stop left
Day clip 8	morning, urban	1,040	2,025	8	1min 4s	Go, go left, stop, stop left
Day clip 9	morning, urban	960	1,264	4	59s	Go, go left, warning left, stop, stop left
Day clip 10	morning, urban	48	109	4	3s	Go, stop
Day clip 11	morning, urban	1,052	1,268	6	1min 5s	Go, stop
Day clip 12	morning, urban	152	229	3	9s	Go
Day clip 13	evening, urban	693	873	8	43s	Go, warning, stop
Night clip 1	night, urban	591	1,885	8	36s	Go
Night clip 2	night, urban	2,299	4,205	25	2min 24s	Go, go left, warning, stop, stop left
Night clip 3	night, urban	1,051	1,476	14	1min 6s	Go, go left, warning left, stop, stop left
Night clip 4	night, urban	1,104	2,538	9	1min 9s	Go, warning, stop
Night clip 5	night, urban	1,453	3,242	19	1min 31s	Go, go left, warning, stop, stop left
		46,418	112,971	290	44min 24s	

`//cvrr.ucsd.edu/LISA/datasets.html` for educational, research, and non-profit purposes.

5. Evaluation

A wide variety of approaches and metrics have been used to evaluate detector performance. A standardized methodology would make comparison more straightforward. For evaluations on the LISA Traffic Light database we suggest using: precision and recall, which are defined in equation (1) and (2). TP, FP, and FN are abbreviations for true positives, false positives and false negatives. The TPs, FPs and FNs should be evaluated on a per frame basis.

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

Precision is the ratio of correct TL detections to the total number of detections made by the detector.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

Recall is the ratio of correct TL detections to the actual number of TLs.

For presenting and evaluating the overall system performance, we suggest generating a precision-recall curve and using the area-under-curve (AUC) as performance measure. A high AUC indicates good performance, an AUC of 100% indicates a perfect system performance for the testset. An example of a precision-recall curve is seen in Figure 4.

The Pascal overlap criterion defined in equation (3) is used to determine TPs:

$$a_0 = \frac{\text{area}(B_d \cap B_{gt})}{\text{area}(B_d \cup B_{gt})} \geq 0.5 \tag{3}$$

a_0 denotes the *overlap ratio* between the detected bound-

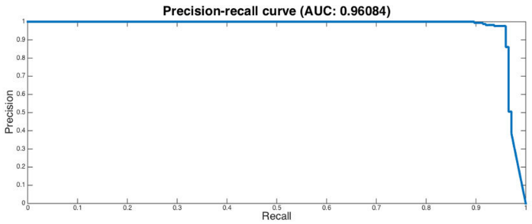


Figure 4: Example of a precision-recall curve

ing box B_d and the ground truth bounding box B_{gt}. $(B_d \cap B_{gt})$ denotes the intersection of the detected and ground truth bounding boxes, and $(B_d \cup B_{gt})$ denotes their union.

We propose that future DAS focused TLR systems are evaluated on a frame-by-frame basis. Furthermore, a proposed system's accuracy should be calculated using AUC for a PR curve. The proposed evaluation terms are listed below:

- True positives are defined according to equation (3).
- Precision, as seen in equation (1).
- Recall, as seen in equation (2).
- Area-under-curve for Precision-Recall curves.

6. Concluding Remarks

We have presented an overview of four state of the art approaches to traffic light detection, which have been published in recent papers on traffic light recognition (TLR). None of the examined TL detection approaches rely on machine learning. This raises the question of whether learning based approaches would be able to outperform these heuristic model based detectors on a challenging dataset such as the one presented in this paper. Because the systems are evaluated using different methodology and on very different datasets it is not clear which approach is the best. The approach proposed in [7] has the major advantage of using prior maps, which significantly ease the detection problem. Only one public database with traffic lights (TLs) is currently available and it is not widely used. We therefore contribute with the LISA Traffic Light Database, which contains TLs captured using a stereo camera on roads in San Diego, USA under varying conditions. The database is supposed to enable comparable evaluation on a large and varied dataset, and provides the possibility of including stereo vision for improving TLR. Along with the dataset we propose a standardized method for evaluating TLR systems, which should enable easy comparison between future detectors. The dataset will eventually be included in the *VIVA Challenge* [12].

References

[1] N. Buch, S. Velastin, and J. Orwell. A review of computer vision techniques for the analysis of urban traffic. *Intelligent Transportation Systems, IEEE*, 12:920–939, Sept 2011.

[2] R. Charette and F. Nashashibi. Traffic light recognition using image processing compared to learning processes. In *Intelligent Robots and Systems*, pages 333–338. IEEE, 2009.

[3] 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*, 7(12):6919–6934, 2011.

[4] R. Danescu, F. Oniga, and S. Nedevschi. Modeling and tracking the driving environment with a particle-based occupancy grid. *Intelligent Transportation Systems, IEEE*, 12:1331–1342, 2011.

[5] R. de Charette and F. Nashashibi. Real time visual traffic lights recognition based on spot light detection and adaptive traffic lights templates. In *Intelligent Vehicles Symposium, IEEE*, pages 358–363, 2009.

[6] M. Diaz-Cabrera, P. Cerri, and P. Medici. Robust real-time traffic light detection and distance estimation using a single camera. *Expert Systems with Applications*, pages 3911–3923, 2014.

[7] N. Fairfield and C. Urmson. Traffic light mapping and detection. In *Proceedings of ICRA 2011*, 2011.

[8] Federal Highway Administration. Manual on uniform traffic control devices, 2015.

[9] U. Franke, D. Pfeiffer, C. Rabe, C. Knoeppel, M. Enzweiler, F. Stein, and R. Herrtwich. Making bertha see. In *Computer Vision Workshops (ICCVW), IEEE*, pages 214–221, 2013.

[10] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Computer Vision and Pattern Recognition*, 2012.

[11] H.-K. Kim, Y.-N. Shin, S.-g. Kuk, J. H. Park, and H.-Y. Jung. Night-time traffic light detection based on svm with geometric moment features. *World Academy of Science, Engineering and Technology 76th*, pages 571–574, 2013.

[12] Laboratory for Intelligent and Safe Automobiles UC San Diego. Vision for intelligent vehicles and applications (viva) challenge, 2015. `http://cvrr.ucsd.edu/vivachallenge/`.

[13] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund. Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey. *Intelligent Transportation Systems, IEEE*, 13:1484–1497, 2012.

[14] Robotics Centre of Mines ParisTech. Traffic lights recognition (tlr) public benchmarks, 2015.

[15] A. Ruta, Y. Li, and X. Liu. Towards real-time traffic sign recognition by class-specific discriminative features. 2009.

[16] S. Sivaraman and M. M. Trivedi. A review of recent developments in vision-based vehicle detection. In *Intelligent Vehicles Symposium*, pages 310–315, 2013.

[17] G. Trehard, E. Pollard, B. Bradai, and F. Nashashibi. Tracking both pose and status of a traffic light via an interacting multiple model filter. In *Information Fusion*. IEEE, 2014.

[18] United Nations. Vienna convention on road signs and signals, 2006.

[19] Y. Zhang, J. Xue, G. Zhang, Y. Zhang, and N. Zheng. A multi-feature fusion based traffic light recognition algorithm for intelligent vehicles. In *Control Conference (CCC)*, pages 4924–4929. IEEE, 2014.