Vision for Looking at Traffic Lights

Issues, Survey, and Perspectives

Jensen, Morten Bornø; Philipsen, Mark Philip; Møgelmose, Andreas; Moeslund, Thomas B.; Trivedi, Mohan M.

Published in:
I E E E Transactions on Intelligent Transportation Systems

DOI (link to publication from Publisher):
10.1109/TITS.2015.2509509

Publication date:
2016

Document Version
Accepted author manuscript, peer reviewed version

Citation for published version (APA):
Jensen, M. B., Philipsen, M. P., Møgelmose, A., Moeslund, T. B., & Trivedi, M. M. (2016). Vision for Looking at Traffic Lights: Issues, Survey, and Perspectives. *I E E E Transactions on Intelligent Transportation Systems*, *17*(7), 1800-1815. https://doi.org/10.1109/TITS.2015.2509509

Vision for Looking at Traffic Lights: Issues, Survey, and Perspectives

Morten B. Jensen, Mark P. Philipsen, Andreas Møgelmose, Thomas B. Moeslund, and Mohan M. Trivedi

***Abstract*—This paper presents the challenges that researchers must overcome in traffic light recognition (TLR) research and provides an overview of ongoing work. The aim is to elucidate which areas have been thoroughly researched and which have not, thereby uncovering opportunities for further improvement. An overview of the applied methods and noteworthy contributions from a wide range of recent papers is presented, along with the corresponding evaluation results. The evaluation of TLR systems is studied and discussed in depth, and we propose a common evaluation procedure, which will strengthen evaluation and ease comparison. To provide a shared basis for comparing TLR systems, we publish an extensive public dataset based on footage from US roads. The dataset contains annotated video sequences, captured under varying light and weather conditions using a stereo camera. The dataset, with it's variety, size, and continuous sequences should challenge current and future TLR systems.**

***Index Terms*—Traffic light recognition, traffic signals, object detection, computer vision, machine learning, intelligent transportation system, active safety, driver assistance systems**

I. Introduction

THE efficiency of transportation systems fundamentally affect the mobility of the workforce, the environment, and energy consumption, which in turn dictates foreign policy. Since transportation is a major part of people's lives, their health and well-being is directly related to it's efficiency, safety, and cleanliness. Many future improvements to transportation systems will come from innovations in sensing, communication, and processing [1], [2].

The automobile revolution in the early 20th century led to a massive increase in road transportation, and the contemporary road network was incapable of handling the rapid increase in traffic load. To allow for efficient and safe transportation, traffic control devices (TCD) were developed to guide, regulate, and warn drivers. TCDs are infrastructure elements that communicate to drivers, e.g. signs, signaling lights and pavement markings [3]. Figure 1 shows an illustration of a road scene with some of the many TCDs.

TCDs are especially important in complex settings such as intersections, where a lot of information must be communicated. Informing drivers is a balance between providing sufficient information while avoiding to burden and distract drivers excessively. A driver's ability to obtain information from TCDs is limited by the amount of information and the time available to comprehend the information. High speed and overwhelming amounts of information may hence lead to errors from oversights and stress [3]. For TCDs to function properly, all road users are required to abide, otherwise dangerous situations occur. Drivers sometimes purposely disregarded TCDs. One study shows that more than 1/3 of Americans admit to having purposefully run a red light during the past month [4]. In many cases, failure to comply is unintentional and caused by e.g. speeding to make it through the intersection in time, aggressive driving by closely following the car in front, distraction [5], misunderstandings, or faulty TCDs.

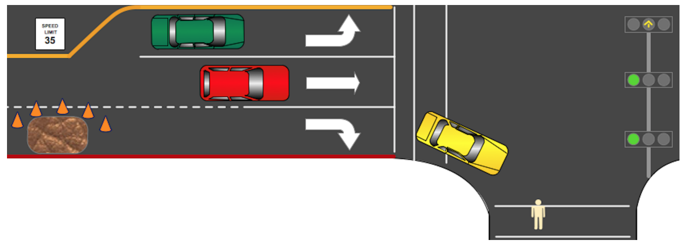


Fig. 1: Traffic control devices for safe and efficient traffic flow.

The complex task of driving is easy, most of the time, since many driving sub-tasks are automated. Effortless driving results in unfocused drivers to whom critical events can be perceived with an added delay. Stressful driving where the driver is very focused and attentive can delayed reaction time, because of fatigue, and mental overload [6].

Widespread autonomous driving lies years in the future, in the meantime lives can be saved by having driver assistance systems (DAS) monitor the environment and warn or intervene in critical situations. For DAS to best support the driver, it must attempt to make up for the driver's deficiencies. An example of a driver deficiency is noticing and recognizing TCDs. Studies show that drivers notice some TCDs better than other; speed limit signs are almost always noticed, while pedestrian crossings signs are mostly overlooked [7].

For all parts of DAS, the urban environment possesses a wealth of challenges, especially to systems that rely on computer vision. An important challenge is recognizing TLs at intersections. In 2012, 683 people died and 133,000 people were injured in crashes that involved red light running in the USA [8]. Ideally, TLs should be able to communicate both visually and using infrastructure to vehicle (I2V) by

Manuscript received May 6, 2015; revised September 9, 2015 and November 26, 2015; accepted December 10, 2015; Date of publication ?; The Editor-in-Chief for this paper was Fei-Yue Wang

M.B. Jensen, M.P. Philipsen and A. Møgelmose are with the Computer Vision and Robotics Research Laboratory, University of California, San Diego, La Jolla, CA 92093-0434, USA and the Visual Analysis of People Laboratory, Aalborg University, 9000 Aalborg, Denmark.

T. B. Moeslund is with the Visual Analysis of People Laboratory, Aalborg University, 9000 Aalborg, Denmark.

M.M. Trivedi is with the Computer Vision and Robotics Research Laboratory, University of California, San Diego, La Jolla, CA 92093-0434, USA.

means of radio communication. Introducing I2V on a large scale requires substantial investments in infrastructure, which are unlikely in the near future. Intersections are some of the most complex challenges that drivers encounter, making visual recognition of TLs an integral part of DAS. The yellow light dilemma is one example where DAS can support drivers. When entering an intersection with a yellow TL, the driver must make a decision of whether to stop or to keep going and cross the intersection. The interval where this decision is difficult for most people is in the range of 2.5-5.5 seconds before entering the intersection [9]. Outside this interval the decision is typically quite clear. The reaction times of drivers is longest in the center of the interval, where the decision is the most difficult. Figure 2 shows two scenarios where information from different sensors and intelligent systems can be combined and provide driver assistance.

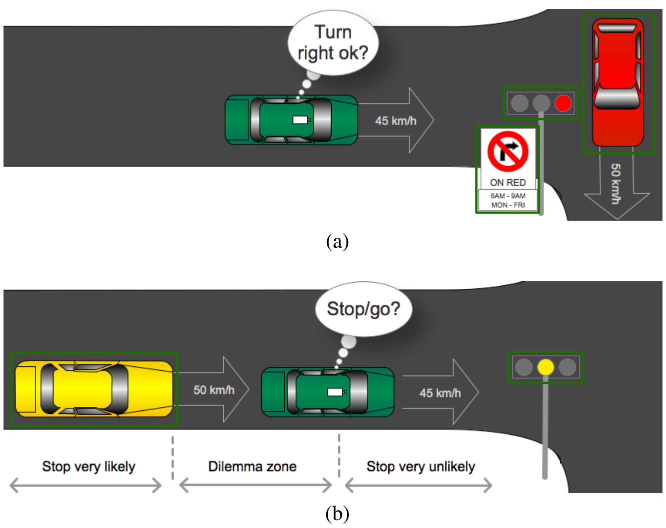


Fig. 2: Fused DAS system in intersection scenarios. (a) Turn right on red assistance. (b) Dilemma zone assistance.

Until now no comprehensive survey of traffic light recognition (TLR) research has been published. The sub-problem of traffic light detection has been survey in [10] and in [11] we presented an introductory overview of ongoing work on traffic light detection along with the LISA Traffic Light Dataset. Most published TLR systems are evaluated on datasets which are unavailable to the public. This makes comparison between existing methods and new contributions difficult. The contributions made in this survey paper are thus, fourfold:

1) Clarifying challenges facing TLR systems.
2) Overview of current methods in TLR research.
3) Common evaluation procedure for TLR systems.
4) High resolution, annotated, stereo video dataset.

The paper is organized as follows: Section II touches on computer vision areas similar to TLR. In section III, TL appearance is discussed along with common challenges facing TLR systems. Section IV presents the typical TLR system. Recent work is examined in section V. In section VI evaluation of TLR systems is reviewed, and a common procedure is proposed. Section VII presents the LISA TL Dataset. In section VIII, experiences, and future possibilities are discussed. Section IX rounds of with the findings made through out the survey.

II. Related Computer Vision Challenges

Before delving into the research made in TLR, it is interesting to examine the challenges, methods, and experiences from related computer vision problems which in many cases will be similar. Related computer vision problems would be: traffic sign recognition, taillight, headlight, and lane detection.

Detection of traffic signs is challenging when subject to varying lighting, viewpoints, and weather conditions. All of these issues suggest that relying solely on color is problematic, therefore shape information is useful for sign detection. An example of the use of shape information is seen in [12]. Relying on shape can also be challenging with both traffic signs and TLs, as the angle between the ego-vehicle and the sign or TL will impact the perceived shape of the object, resulting in a new shape variation. Developing robust vision based DAS that works under changing lighting, at varying distances, and under mixed weather conditions is a difficult task as stated in [13], which mentions that cross-over procedures for handling environmental transitions should be investigated. Following the 2012 survey on traffic sign recognition [7], the focus has shifted entirely to learning-based traffic sign detectors and the problem is considered solved on a subset of signs [14], [15]. The same paradigm shift has not yet materialized in TLR.

Most vehicle detection and brake light detection at night utilize monocular cameras and rely on the symmetry of tail and head lights for detecting vehicles as seen in [16], [17], [18], [19], [20], [21]. [22] detects head and tail lights using cues from lane detection, with the purpose of automatic switching between high-beam and low-beam. Similarly, cues from lane detection are important additions to TLR systems in order to determine the relevance of TLs. A recent paper on lane detection is [23], where a context aware framework for lane detection is introduced, this can significantly reduce the required computational demand by scaling the detection algorithm based of the state of the ego-vehicle and the road context. The same paper references several comprehensive surveys on lane estimation techniques, one being [24], where work done across multiple modalities is reviewed. In [25], [26] the gaze and attention of the driver is determined. This is essential information for DAS, since it can be used to determine if the driver should be notified as e.g. in [27] where the driver is alerted and safety systems are engaged if the driver is inattentive for a prolonged period of time.

III. Traffic Lights: Conventions, Structure, Dynamics, and Challenges

TLs regulate the traffic flow, by informing drivers about the right of way. Right of way is given in a manner which minimizes conflicts between vehicles and pedestrians traveling incompatible paths through the intersection. TLs are by design made to stand out and be easily visible. Their primary components are bright colored lamps, usually circle or arrow shaped. The lamps are surrounded by a uniformly colored container. The most common TL configuration is the red-yellow-green light, where each state indicates whether a driver should stop, be prepared to stop, or keep driving. A variety of other TLs have been created as a result of complex intersections. Figure 3 shows some of the allowed vertical configurations of TLs in California.

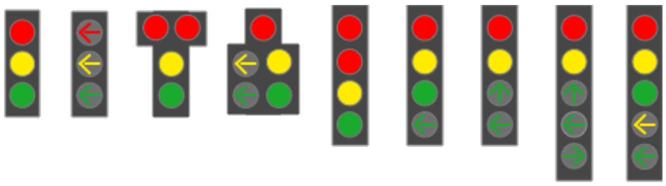
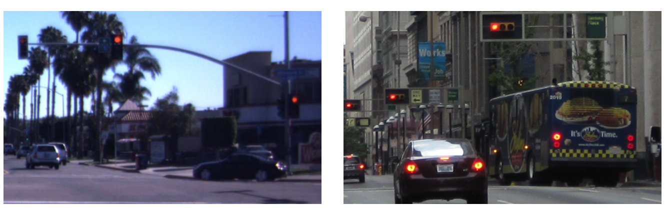


Fig. 3: Examples of vertical TLs found in California. [28]

The orientation, color, size, and shape of the container will vary country to country and even city to city. An example of differently oriented and colored TLs within the USA is seen in Figure 4. There are two methods for mounting TLs, suspended and supported, this is evident in Figure 4(a). The supported variety has proven the most difficult for existing TLR systems, as discussed in subsection III-A.



(a) (b)

Fig. 4: (a) San Diego, California. (b) Cincinnati, Ohio.

Besides the various configurations of TLs, the state sequence is an important characteristic of a TL. An example of a state sequence for the basic red-yellow-green light is shown in Figure 5.

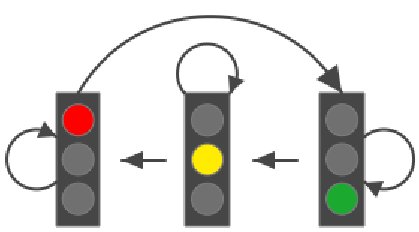
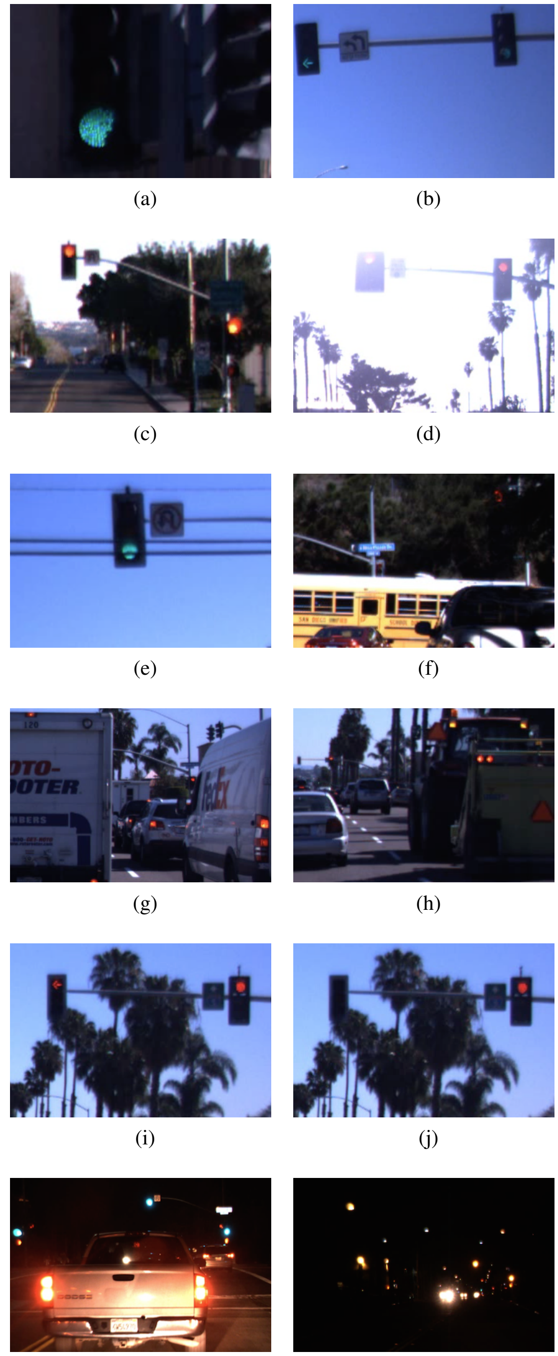


Fig. 5: Basic TL sequence for states: green, yellow, and red.

For increasing road safety and making it easier for drivers when driving across states, TLs in USA are regulated by the Federal Highway Administration in the *Manual on Uniform Traffic Control Devices* [29]. Most countries in Europe have signed the *Vienna Convention on Road Signs and Signals* [30], requiring TLs to meet a common international standard.

A. Challenges in recognizing traffic lights

Although TLs are made to be easily recognizable, influences from the environment and e.g. sub-optimal placement can



(a) (b)

(c) (d)

(e) (f)

(g) (h)

(i) (j)

(k) (l)

Fig. 6: Examples of frames from the collected dataset.

make successful detection and recognition difficult, if not impossible. Issues include:

- Color tone shifting and halo disturbances [31] e.g. because of atmospheric conditions of influences from other light sources . Figures 6(c), 6(d), 6(k) and 6(l).
- Occlusion and partial occlusion because of other objects or oblique viewing angles [31]. Especially a problem with supported TLs [32], [33], [34]. Figures 6(e) to 6(g).
- Incomplete shapes because of malfunctioning [31] or dirty lights. Figure 6(a).

- False positives from, brake lights, reflections, billboards [35], [36], and pedestrian crossings lights. Figure 6(h).
- Synchronization issues between the camera's shutter speed and TL LED's duty cycle. Figures 6(i) and 6(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 containing a lit TL lamp. Saturation is another aspect that can influence the appearance of the lights. With transition between day and night, the camera parameters must be adjusted to let the optimal amount of light in and avoid under or over-saturation. [37] 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 image.

IV. Traffic Light Recognition for Driver Assistance Systems

Computer vision problems like TLR can be divided into three sub problems: detection, classification, and tracking. The typical flow of such a system is illustrated in Figure 7.

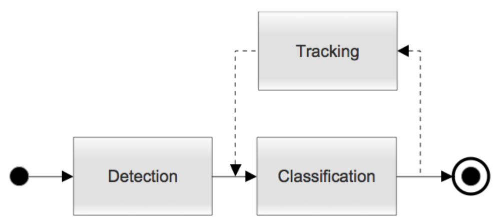


Fig. 7: Breakdown of a computer vision based TLR system.

[7] presents a similar breakdown for traffic sign recognition. The detection and classification stages are executed sequentially on each frame, whereas the tracking stage feeds back spatial and temporal information between frames. The detection problem is concerned with locating TL candidates. Classification is done based on features extracted from the detected candidates. Tracking uses information about location and TL state when tracking TLs through a sequence of frames. A TLR system that addresses the mentioned problems can therefore be broken into 4 stages: detection, feature extraction, classification, and tracking.

A. Challenges in recognizing traffic lights for DAS

When several TLs are simultaneously visible to the driver, each possibly in different light states, the assistance system must be able to determine which TL is relevant to the driver, a task that can be difficult, even to a human. Figure 8 shows an example of a complex traffic scene where three upcoming intersections are all visible at the same time. One of the intersections contains turn lanes that are accompanied by their own independent TLs. This represents a major challenge for TLR systems in relation to DAS, in determining whether a TL is relevant to the ego-vehicle.

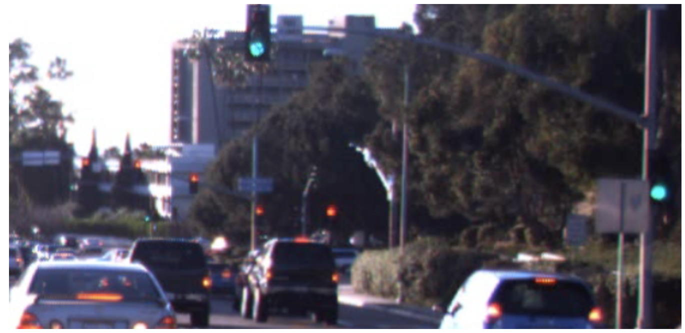


Fig. 8: Complex traffic scene with multiple visible intersections and turn lanes, each with their associated TLs.

The relevance of a TL is closely connected to it's placement in relation to the ego-vehicle. Information about the location and direction of the ego-vehicle must therefore be matched with the locations of the TLs. The most advanced system for solving this problem is seen in [35], where a guess is made based on the intersection width and the estimated orientation of the TLs. An alternative and less dynamic approach is used in [32], where the route is recorded beforehand and relevant TLs are manually annotated offline. Features are extracted in the annotated regions, and the system is then able to recognize the relevant TLs on that specific route.

A TLR system for DAS must communicate the gathered information to the driver, preferably in a way that is non-intrusive and adds as little as possible to the cognitive load of the driver. Information about the driver's attention can be used to activate a given safety system in case the driver is inattentive or to determine whether a driver has noticed a specific object and should be made aware of it. Hence, fusion of data from looking-in and looking-out sensors can be used [38]. In [39] a large set of looking-in activities: head pose estimation, hand and foot tracking; and looking-out activities: vehicle parameters, lane and road geometry analysis, and surround vehicle trajectories, are fused together to predict driver behavior. The presentation aspect of DAS is outside the scope of this paper.

V. Traffic Light Recognition: State-of-the-Art

In this section, we present an overview of methods used in each stage of the pipeline for existing TLR systems. The pipeline is divided into four stages; detection, feature extraction, classification, and tracking. This breakdown was described in chapter IV. In addition to the breakdown in the four stages, papers are also presented with their applied color spaces. This is done since the choice of color space is central to TLR and emphasized in some work e.g. [37]. Table I contains an overview of the applied methods for all the papers from academic institutions. Table II shows a similar overview for industry papers. It should be noted that some of the papers presents more than one approach, whereof only the best performing is listed. Since some of the papers focus on only parts of the problem, and in a few cases it is not apparent what methods were used, some fields are left empty.

The paper overview covers papers from 2009-2015, with a single exception of an important paper [40] from 2004, which forms the basis for the more recent paper [32].

A. Color space

As color is a major characteristic of TLs, it is used extensively in most TLR research. It is primarily used for finding region of interest (ROI) and classifying TL states. The variety in color spaces is large, but the RGB color space is often discussed, as it usually is the color space in which input images are represented. Because color and intensity information are mixed in all the channels of the RGB color space, the information is usually translated to another color space for processing. [36], [41] are the only studies, where RGB is the primary color space. The same author group also utilizes the YCbCr color space in their earliest paper [42]. [43] uses both RGB and YCbCr, in two separate stages, RGB is used for localizing the box of the TL, whereas YCbCr is used for detecting the arrow TL. In [44], CIELab is used when extracting features for TL detection, subsequent they also employ RGB for extracting features for TL classification. Normalized RGB has seen used by it self, as in [45] and combined with RGB in [37], [46], [47]. [48], [49], [34], [33] use grayscale for initial segmentation in the form of spot light detection. Whereas [49], [34], [33], rely purely on grayscale, hence, their systems must function using only intensity and shape information. Other works that use grayscale, are [50], where normalized grayscale is used in addition to CIELab and HSV and [51] where grayscale is used for finding candidates, before determining their state using the HSV color space. The HSV and HSI color spaces are well represented by their use in [52], [53], [54], [48], [55], [51], and [56], [57], respectively. It is noteworthy that [53], demonstrates that the hue distribution is much narrower if a low exposure is used when capturing frames. The narrower distribution greatly helps in later segmentation of the frames by limiting color saturation. For each low exposure frame they also capture one with normal exposure to maintain a balanced intensity. [58] uses the HLS color space for determining the state of found TLs amd [31] uses IHLS which is an modification of HLS, which separates chromatic and achromatic objects well. [59], [60] uses the LUV color space for extracting color features.

There is no clear tendency towards the use of one particular color space, but color spaces where color and intensity information is separate are clearly preferred. In some recent work such as [50], [37] and to some degree [61], researches have begun combining channels from multiple color spaces in order to get optimal separation of TL colors. In [62], [63], the CieLab color space is used to create a new channel by multiplying the lightness with the sum of the a and b channels.

B. Detection

TLR systems usually look for a selection of TL components. Figure 9 shows an illustration of the various TL components. The structural relationship between the components is in some cases also used.

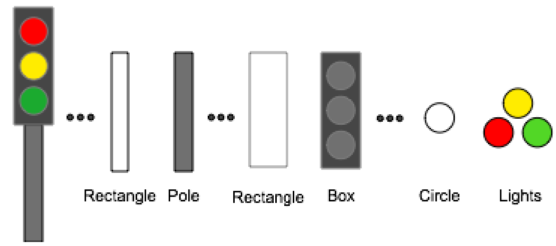


Fig. 9: Supported TL along with it's different components.

Detection approaches can be categorized as either learning-based or model-based, the latter is currently the far most widely used. It relies on heuristically determined models using shape, color and intensity.

Model-based: A simple and widely used color model consists of color density thresholds for each TL lamp color, this is seen in [36], [42], [47], [53], [63], [48], [45], [35], [55], [57], [31], [41]. Detectors based on this kind of simple color model are especially in danger of suffering from overfitting to the specific training set. [32], [40] present several detectors, the best performing being a Gaussian pixel classifier, created from thousands of manually annotated images. In [37], [46], fuzzy clustering is introduced to generate unique representations of a given color. Opposite to regular clustering, sometimes called hard clustering, data points in fuzzy clustering can belong to more than one cluster; the association degree to clusters can vary, meaning that a data point can have a higher probability of belonging to one cluster than another [64]. Common for all detectors relying on color distributions is their sensitivity to disturbances in color which can be caused by many of the effects described subsection III-A. Spotlight detection using the white top hat operation is a popular detection approach which is robust to disturbances in color. It is usually based on greyscale images, as seen in [48], [49], [34], [33]. In [52], the V channel from the HSV color space is used with the same effect.

Shape models are either used an alternative to the dominating color models or as a filtering step after the color based segmentation In [41] the Hough transform is used on an edge map from a Laplacian edge detection filter. The contribution of [45] is a modified version of the circular Hough transform, which looks for filled circles, and outperforms the conventional circular Hough transform. In [61] they improve this idea further by also looking for other circles around active lights. [31] first apply a Laplacian filter that extracts a clear boundary, disregarding halo effects, before looking for approximate ellipses in the canny edge pixels around candidate BLOBs. In [63], [62], fast radial symmetry is used for finding circles, followed by local maximum and minimum for finding the exact parameters of the given circle. [51] finds object boundaries using morphological operations and thresholding, the borders are topologically analyzed and TL candidate rectangles are located.

Most of the reviewed work applies BLOB analysis in various degrees for noise removal or for calculating BLOB properties. [49], [34], [36], [33], [42], [43], [48], [56], [57] removes noise by looking at a selection of BLOB features, from relative position of elements such as circles, squares, rectangles, spots, containers, to size, aspect ratio, shape, etc.. An example is [48] where BLOB size, aspect ratio, circular shape, holed regions, and the estimated height in world coordinates is used. [57] employs region growing using seed points from their found BLOBs to perform a border property check between found BLOBs and their corresponding grown regions. Other BLOB analysis includes doing bounding box analysis as in [37], [46], [47], [52], [53], where the goal is to locate the TL box such that the state within it can be estimated. Instead of finding shape from segmented BLOBs, [45], [51] applies a Sobel Kernel to get an edge map and applies Hough transform in order to find either circular shapes or boxes. Using BLOB analysis to filter out false TL candidates is very dependent on the quality of the BLOB segmentation. In many cases BLOBs from actual TLs will appear vastly different from a TL BLOB under ideal conditions.

Learning-based: An early attempt at learning based detection was seen in [32], [40] where a cascading classifier based on Haar features was tested. It was, however, outperformed by their Gaussian color classifier. Recently, three papers that employ some more successful learning-based detectors have been published. [58] is combining occurrence priors from a probabilistic prior map and detection scores based on SVM classification of Histogram of Oriented Gradients (HoG) features to detect TLs. [44] detects TL pixels by extracting features from color, texture, pixel location and normalized voting histograms and classify them using a JointBoost classifier. In [59], [60] features are extracted as summed blocks of pixels in 10 channels created from transformations of the original RGB frames. The extracted features are classified using depth 2 learning trees in [59]. In [60] detector performance is improved by increasing the learning tree depth to 4 and extending the scale space. Common for the learning-based detectors is their requirement for lots of data and computation. Their advantages are a higher robustness to variation and less tendency to overfitting because of the substantial amount of data used in the training process.

Auxiliary detection: In [41] GPS position is used to activate the detection system when approaching preannotated intersections. [32], [40], [54], [35], [52], [58] are taking this further by improving detection by including maps containing much more precise prior knowledge of TL locations. The maps are created using accurate GPS measurements and manual annotation of TL position. In [54] they store hue and saturation histograms of each TL during the initial TL mapping. This helps with handling differences in the light emitting properties of individual TLs. In [35] the possible states of the individual TLs is also annotated to further reduce false positives. Relying on prior maps can be a big help to visual TL detection. The maps increase the certainty in TL candidates from the detector which makes it easier to reject false candidates. In cases where a TL has not been mapped e.g. due to roadwork, a high reliance on maps might lead to critical missed detections.

Generally, the first step of model-based detection is segmentation of ROI by using either, clustering, distributions, or thresholds. This is followed by either looking for circular objects using Hough transform or fast radial symmetry, or BLOB analysis to filter TL candidates. With learning-based detectors all of this is achieved by the classification of numerous features. Prior knowledge of the route, geographical and temporal information can drastically reduce the ROI, hence reduce the computational requirements and the number of bad candidates.

C. Feature Extraction

Color is a widely used feature for classification. This is seen in [37], [46], [47], [32], [40], [51], [52], [50], [35], [56], [45], [61], where color densities from segmented areas are used. In [54], [48] the features are based on HSV histograms. Besides color, shape and structure are widely used characteristics from TLs. Shape information includes a wide variety of features, such as aspect ratio, size, and area. Structural information is the relative positioning of the components of TLs. A TL lamp and the surrounding container is, in many cases, easily distinguishable from the background, making shape, and structural information popular features. [57] uses color template matching. Shape information is combined with structural information in [49], [34], [33], and also color features as seen in [37], [46], [47], [50], [45], [61]. In [52], [41], color, shape, and structural information is used as features. In [36], a mix of BLOB width, height, center coordinate, area, extent, sum of pixels, brightness moment, and geometric moment is used as features for their SVM classifier. More advanced feature descriptors are seen in [53], where edge information in the form of HoG, are used as features for image regions containing TL containers. [43] uses 2D Gabor Wavelet features and [55], [42] uses Haar features. [44] extract 21 geometric and color features from TL candidate regions. [31] relies on spatial texture layout for classification, specifically they calculate a Local Binary Pattern (LBP) histogram for the TL as well as for five equally sized regions in each color channel, before creating a feature vector consisting of the concatenated LBP histograms.

Systems relying either color, shape, or structural features will be challenged in varying conditions of the real world. By using multiple types of features containing different types information increase robustness.

D. Classification

For classification of TL candidates [52] utilizes a fusion between scores from structure, shape, color, and geolocation information, which help determine whether a TL should exist at that location. [51] simply estimate the state to be the winner of a majority count on the number of pixels within empirically determined thresholds. [56] decides on a TL state for the entire segmented frame based on a contradiction scheme that selects the optimal light from TL position and size. [58] classifies TL candidates by subdividing them vertically in three using the color distribution. [43] focus on classifying the arrow type of their TL candidates, this is done by nearest neighbor classification of Gabor image features which are reduced by 2D independent component analysis. [53] classifies the TLs by using HoG features from the TL container and SVM. In [32], [40], a neural network is used for determining the state of the found TLs. [57] applies template matching by normalized cross correlation. [34], [33], [49] use adaptive template matching. [48] uses SVM for classification based on HSV histograms. [42], [55] use cascading classifiers based on Haar features. [34] compares their proposed adaptive template matching system with the learning-based AdaBoost cascade Haar feature classifier. Their model based approach proved to substantially outperform their learning based approach. [44] classify their 21 element feature vector using a JointBoost classifier. [31] applies SVM to classify LBP feature vectors in order to determine the state of a TL from it's spatial texture layout.

Successful classification rely heavily on the quality of the features. The majority of papers apply a classifier to the extracted features and find the best match by comparison with a selection of trained TL states. The remaining papers classify TLs based on heuristics. Classification based on e.g. heuristically determined thresholds, is vulnerable to many of the variations found under real world use. The machine learning-based approaches train a model based on training samples, which requires large amount of data with large variation in order to obtain robustness.

E. Tracking

Tracking is commonly used for noise reduction by suppressing false positives and handling dropouts due to e.g. occlusion. It is evident in Table I and II that approximately half of the presented approaches apply some form of tracking.

Temporal tracking is used to examine previous frames and determine whether a candidate has been found in the same area earlier and whether it has the same properties as a given candidate in the current frame. This is a simple and straightforward approach used in [46], [47]. [56] reduced false detections by a third by using a temporal decision scheme that makes a final classification based on temporal consistency. A similar approach is seen in [44], where a TL has to be detected in three consecutive frames before it is accepted. It is evident their results that including this type of tracking led to an increase of 12.16% in overall precision, while costing 6.27% in overall recall. [48] employs multiple target temporal tracking using predictions of the location of TL from the speed of the ego vehicle. This allows for validation of state changes and smoothed recognition confidence. Additionally, they modify top hat kernel size, saturation, and intensity thresholds when TLs are about to disappear from the field of view. This enables recognition in a greater distance interval. Before reaching the final state verdict, [48] inputs the detected state from their classifier into a range of HMMs, one for each possible type of TL and one for a non-TL objects. The model which best fits the detected sequence of states is then selected as the final estimated state. [51] also employs HMM, although only for a single TL type. [37] estimates the distance to TLs using inverse perspective mapping and tests both a Kalman filter and a particle filter for tracking the relative movement between vehicle and TLs. TLs are then filtered based on their consistency in position and color. In [49], an Interacting Multiple Model filter is used for keeping track of both state and position of a given TL. The prediction in the model is using Kalman filters to keep track of the state and the position in time. For establishing the state, a Markov chain with weighted probabilities is used for finding the current state based on posterior states, originally introduced in [65]. [54] uses prior maps and a histogram filter to adjust the localization mismatch between predict and actual TL area.

The correlation tracking used in [52], [32], [40], relies on the fact that a given detected TL's state is unlikely to shift sporadically in a sequence of frames. E.g., when a series of red states are detected, it is most likely that the state in the upcoming frame will be red again and the appearance will therefore be approximately the same. [55] use CAMSHIFT tracking of candidates across frames based on their appearance.

Tracking is mostly used to filter out noise and handle lone failed detections, caused by e.g. occlusion. In most of the surveyed papers, tracking consist of a simple temporal consistency check, a few use tracking in a more advanced manner by incorporating prior probabilities. Generally, two types of tracking are used, correlation tracking and point tracking. In many cases correlation tracking rely on the same types of features as the detector and for this reason will be unable to complement the detector when it fails. Point tracking on the other hand can employ temporal information which has a better basis for complementing the detector.

VI. Evaluation

Performance of TLR systems has been evaluated in a wealth of ways throughout the reviewed work, complicating comparisons between competing approaches. Additionally, some papers does not clearly define which evaluation criteria have been used. Evaluation is generally done on a local collection of frames, unavailable to the public. These local datasets are mostly small in size and contain little variation.

A. Performance measures

The most common measures of system performance are: precision, recall, and accuracy. Results from the reviewed TLR systems are therefore, when possible, summarized using these measures. Precision, recall, and accuracy are defined as in [66]. The definitions are shown in equation (1), (2) and (3). TP, FP, FN, TN are abbreviations for true positives, false positives, false negatives, and true negatives.

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

A Precision close to one indicates that all the recognized TL states are in fact correctly recognized.

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

A recall close to one indicates that all the TL states, in a given video sequence, were correctly recognized by the proposed system.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{3}$$

TABLE I: Recent academic studies in TLR. Colors indicate paper group affiliation. Corresponding evaluation results and datasets for each paper are available in Table III. Abbreviations: Connected component analysis (CCA), support vector machine (SVM), hidden Markov model (HMM)

Paper	Year	Color Space(s)	Detection	Features	Classification	Tracking
[37]	2014	RGB, RGB-N	Fuzzy clustering, CCA, BLOB analysis	Color	Color	Kalman filter, particle filter
[46]	2013	RGB, RGB-N	Fuzzy clustering, CCA, BLOB analysis	Color	Color	Temporal filtering
[47]	2012	RGB, RGB-N	Clustering, BLOB analysis	Color	Color	Temporal filtering
[49]	2014	Grayscale	Top-hat spot light detection, BLOB analysis	Shape, structure	Adaptive template matching	Interacting Multiple Model
[34]	2009	Grayscale	Top-hat spot light detection, BLOB analysis	Shape, structure	Adaptive template matching	-
[33]	2009	Grayscale	Top-hat spot light detection, BLOB analysis	Shape, structure	Adaptive template matching	-
[36]	2013	RGB	Color thresholding, BLOB shape filtering	Brightness and geometric moments	SVM	-
[42]	2011	YCbCr	Color thresholding, BLOB shape filtering(width to height ratio, sum of pixels, BLOB area to bounding rectangle ratio)	Haar-like features	Adaptive multi-class classifier trained using AdaBoost	-
[61]	2010	RGB, RGB-N	Pixel clustering, edge map, filled circle Hough transform in neighborhood	Color, shape	Color of best circle	-
[45]	2009	RGB-N	Color thresholding, edge map, filled circle Hough transform	Color, shape	Color of best circle	-
[59]	2015	LUV	Aggregated channel features	-	-	-
[60]	2015	LUV	Aggregated channel features	-	-	-
[44]	2015	RGB, CIELab	Color, texture, pixel location and normalized voting histogram classified with JointBoost	Color, Geometric features	JointBoost	Temporal filtering
[58]	2015	HSL	Probabilistic prior maps and dense HoG	Color	Color distribution	-
[51]	2014	Grayscale, HSV	Topological analysis of edges	Color	Majority pixel count	HMM
[52]	2014	HSV	Top-hat spot light detection, BLOB analysis	Color, shape, structure	Fusion of color, shape, structure scores, and geolocation	Correlation tracking
[53]	2014	HSV	Color thresholding, BLOB analysis	HoG	SVM	-
[50]	2014	Norm. grayscale, CIELab, HSV	Prior knowledge of TL location	Color, shape	Convolutional neural network	-
[63]	2014	CIELab	Color thresholding, radial symmetry, local maximum and minimum, shape filtering	-	-	-
[43]	2012	RGB, YCbCr	BLOB analysis	2D Gabor wavelet	Nearest neighbor	-
[62]	2012	CIELab	Color difference enhancement, neighborhood image filling, radial Symmetry	Color	Color	Spatial-temporal consistency check
[41]	2012	RGB	Color thresholding, edge map with Laplace filter, circle Hough transform, shape filtering	Color, shape, structure	Color	-
[57]	2011	HSI	Color thresholding, dimensionality and border property check	Color	Normalized cross correlation template matching	-
[31]	2011	IHLS	Color thresholding, Laplacian filter for boundary extraction, approximate ellipses based on edge pixels from canny	LBP features	SVM	-
[54]	2011	HSV	Prior knowledge of traffic light location	HS histograms	Histogram back-projection	Histogram filter
[48]	2010	HSV	Top-hat spot light detection, Color thresholding, CCA, BLOB analysis	Concatenated HSV histogram	SVM	HMM and temporal tracking using ego motion
[55]	2010	HSV	Color thresholding, morphological operation	Haar features	AdaBoost trained classifier	CAMSHIFT
[56]	2009	HSI	Gaussian-distributed classifier, BLOB analysis, temporal information	Color	Global contradiction solving scheme	Temporal filtering/decision scheme

TABLE II: Recent studies in TLR from industry. Colors indicate paper group affiliation. Corresponding evaluation results and datasets for each paper are available in Table IV.

Paper	Year	Color Space(s)	Detection	Features	Classification	Tracking
[32]	2013	-	Prior knowledge of TL location, Gaussian-distribution classifier	Color	Neural network	Correlation tracking
[40]	2004	-	Prior knowledge of TL location, Gaussian-distribution classifier	Color	Neural network	Correlation tracking
[35]	2011	-	Prior knowledge of TL location and state sequence	Color, shape	Color and BLOB geometry	Temporal filtering

An accuracy close to one indicates that the system detects all TLs with no false positives. Traditionally, true negatives are also included in the calculated accuracy as follows from equation (3), but true negatives are rarely used in evaluation of TLR systems.

In some cases these performance measures are referred to under different names, e.g. detection rate instead of recall or recognition rate for accuracy. When it is unclear what the used terms are covering, they are published as accuracy in Tables III, IV and V. The criteria for deciding when a TL is recognized and registered as a true positive can also be unclear or vary widely. An example of this is seen in [34], [33], where a TL is registered as a TP if it has been recognized once in the whole series of frames where it appears. We suggest evaluating FPs, TPs on a frame by frame basis, as this gives a more complete representation of a given system's performance.

B. Evaluation overview

In Table III evaluation data specifications are presented along with the stated results from TLR papers originating from academic institutions. Table IV presents the same for TLR research originating from industry. Since a few papers have published results limited to TL detection, these results are presented separately in Table V. When looking at the tables it is clear that the majority of systems are evaluated on local datasets. Many of these datasets are not described sufficiently and consist of as little as 35 ground truth TLs. Taking dataset size and evaluation results in to account, [37] is one of the best performing systems. TLs are detected based on fuzzy clustering, the system has been refined from earlier publications [46], [47] and recognition is supported by adaptive image acquisition and tracking. Another notable system, with impressive results, is presented in [58], where TLs are detected and recognized in an extensive dataset using a probabilistic prior map and detection based on HoG features. [32] presents an autonomous car with TLR which has successfully driven a 100 km route in real traffic. TLs are detect using prior maps and from TL color distributions. The performance of the TLR system is, however, not quantified, atleast not publicly, making it impossible to do direct comparisons to other work.

C. Proposed evaluation methodology

A variety of performance measures can be used for evaluation of TLR systems, examples are recognition rate, detection rate, recall, precision, true positive rate, false positive rate, false positives per frame, confusion matrix, F1-score, etc. No matter which measure is used, it is important to define it clearly. Furthermore, it is important to describe the used datasets in detail in order to make fair assessments possible.

True positive criteria: It should be clear what constitute a true positive. We suggest using the PASCAL overlap criterion introduced in [67]. It is defined as seen in equation (4).

$$a_0 = \frac{\text{area}(B_d \cap B_{gt})}{\text{area}(B_d \cup B_{gt})} \geq 0.5 \tag{4}$$

a_0 denotes the *overlap ratio* between the detected bounding box B_d and the ground truth bounding box B_{gt}.

Describing performance: According to [68] it can be misleading to state performance using overall accuracy on an unevenly distributed dataset. The overall accuracy does not necessarily represent the performance on smaller classes. The distribution of classes in TL datasets is naturally skewed since e.g. the standard stop light is a lot more common than the warning arrow light. When evaluating multi-class classifiers on skewed datasets, confusion matrices provide a great overview of the performance for specific classes. In table VI we present an example of a confusion matrix for a basic TLR system with the stop, warning, and go classes. Dark grey indicates the variables, and brighter grey indicate either ground truth or system classification output. From these numbers, recall, precision, and accuracy seen in the blue fields, can be calculated. The classifications in the confusion matrix for this example were found based on thresholds that provide a high recall at the cost of precision.

TABLE VI: Confusion matrix of 3-class skewed dataset.

			System classification			
			Stop	Warning	Go	Recall
			9210	5102	6194	
Ground Truth	Stop	9703	5.887	160	0	60.67%
	Warning	691	1	406	0	58.76%
	Go	7688	2	0	4.872	63.37%
	Precision		63.92%	7.96%	78.66%	40.71%

The vast majority of papers report performance using the three measures described in subsection VI-A. The problem with these measures is that they only provided a narrow glimpse into the actual performance of the system. By calculating accuracy or precision and recall for a large number of thresholds and plotting the resulting curves, it is possible to observe the performance over the full spectrum of the system. The two most common types of curves are *Receiver Operator Characteristic* (ROC) curves, cost curve, and *Precision-Recall* (PR) curves. In [69], the relationship between ROC curves and PR curves is presented, and it is concluded that when using skewed datasets, the PR curves provides a more informative picture of a system's performance. Both [69] and [68] mention *Area-Under-Curve* (AUC) as an alternative to the traditional measures for comparing algorithms. AUC can describe performance in the full spectrum using a single number. When calculating the AUC it is important to keep in mind that using few thresholds as basis for generating the curve may lead to a poor representation of the systems performance. Ideally the number of thresholds should match the number of different scores outputted by the TLR system.

Figure 10 shows PR curves for the same TLR example used in table VI. The optimal threshold for a given system can be easily determined using the PR curves. Additionally, the AUC will reflect the dramatic drop in precision of the stoplight classifier, a single precision, or accuracy measure on the other hand could not.

The proposed evaluation terms and criteria are listed below:

TABLE III: Evaluation datasets and corresponding results from recent studies in TLR from academia. Colors indicate paper group affiliation. RCMP abbreviates *Robotics Centre of Mines ParisTech* and signifies the use of their dataset, - indicates evaluation on part of the dataset and + indicates the addition of private datasets. Under *Ground Truth*, # of frames indicates # of frames with a minimum of 1 TL.

Paper	**Year**	**Dataset**	**Size [Frames]**	**Ground Truth**	**Resolution [Pixels]**	**Conditions**	**Pr [%]**	**Re [%]**	**Ac [%]**
[37]	2014	Local	75,258	19,083 frames	752x480	Day, night	99.38	98.24	99.39
[46]	2013	Local	16,176	4,600 frames	752x480	Night	98.04	-	-
[47]	2012	Local	27,000	14,000 frames	752x480	Day, night	90.32	-	-
[49]	2014	Local	-	-	-	-	97.6	87.57	97.6
[34]	2009	RCMP+	>11,179	10,339 TLs	640x480	Day	84.5	53.5	-
[33]	2009	RCMP+	>11,179	>9,168 TLs	640x480	Day	95.38	98.41	-
[36]	2013	Local	16,080	12,703 frames	620x480	Night	-	93.53	-
[42]	2011	Local	30,540	16,561 frames	620x480	-	-	93.80	-
[61]	2010	Local	35	35 TLs	-	-	-	-	89.0
[45]	2009	Local	30	30 TLs	-	-	-	-	86.67
[44]	2015	Local, RCMP-	-	-	640x480+	Day, night	72.83	80.13	-
[58]	2015	Local	-	9,301 frames	6x1024x768	Early morning, afternoon	92.3	99.0	-
[51]	2014	Local	649	446 TLs	648x488	Mixed day-time illumination conditions	99.59	92.19	94.45
[52]	2014	Local	3,767	-	-	Day	-	-	96.07
[53]	2014	Local	-	-	640x480	Mixed day-time illumination conditions	-	-	-
[50]	2014	Local	3,351	3,351 TLs	-	Afternoon, dusk	-	-	97.83
[43]	2012	Local	5,000	-	1392x1040	Mixed day-time illumination conditions	-	-	91.00
[62]	2012	RCMP	11,179	9,168 TLs	640x480	Day	61.22	93.75	-
[41]	2012	Local	7,311	-	-	Day	-	89.9	-
[57]	2011	RCMP-	5,553	5,553 frames	640x480	-	96.95	94.4	-
[31]	2011	Local	714	763 TLs	640x480	Sunny, cloudy, rainy	34.51	94.63	95.01
[54]	2011	Local	-	-	1280x1024	Noon, dusk, night	81.46	77.98	92.85
[48]	2010	Local	-	2,867 TLs	512x384	-	-	-	89.6
[55]	2010	Local	-	-	780x580	-	-	-	-
[56]	2009	Local	6,630	-	640x480	-	-	-	98.81

TABLE IV: Evaluation datasets and corresponding results from recent studies in TLR originating from industry. Colors indicate paper group affiliation. Under *Ground Truth*, # of frames indicates # of frames with a minimum of 1 TL.

Paper	**Year**	**Dataset**	**Size [Frames]**	**Ground Truth**	**Resolution [Pixels]**	**Conditions**	**Pr [%]**	**Re [%]**	**Ac [%]**
[32]	2013	Local	-	-	-	100 km in real world	-	-	-
[40]	2004	Local	-	-	-	-	-	>95	-
[35]	2011	Local	-	1,383 frames	2040x1080	Morning, afternoon, night	99.65	61.94	93.63

TABLE V: Evaluation datasets and results from TL detection papers. Under *Ground Truth*, # of frames indicates # of frames with a minimum of 1 TL.

Paper	**Year**	**Dataset**	**Size [Frames]**	**Ground Truth**	**Resolution [Pixels]**	**Conditions**	**Pr [%]**	**Re [%]**	**Ac [%]**
[59]	2015	LISA TL	14,386	21,421 TLs	1280x580	Mixed day-time illumination conditions	30.1	50.0	-
[60]	2015	LISA TL	11,527	42,718 TLs	1280x580	Night-time	65.2	50.0	-
[58]	2015	Local	-	9,301 frames	6x1024x768	Early morning, afternoon	97.3	99.0	-
[63]	2014	Local	70	142 TLs	240x320	Day	84.93	87.32	-

- True positives are defined according to the PASCAL overlap criterion.
- Precision, as seen in equation (1).
- Recall, as seen in equation (2).
- Area-Under-Curve for Precision-Recall curve.
- Confusion matrix.

VII. Traffic Light Dataset

Extensive and challenging datasets are essential for evaluation and comparison of TLR research. Until now the only publicly available dataset was the TLR benchmark from *LaRA (La Route Automatisée)* at Mines ParisTech, Paris. In this section the new Traffic Light Dataset from *LISA (Laboratory for Intelligent and Safe Automobiles)* at University of California, San Diego, is described in detail. Table VII provides an overview of these two TL datasets.

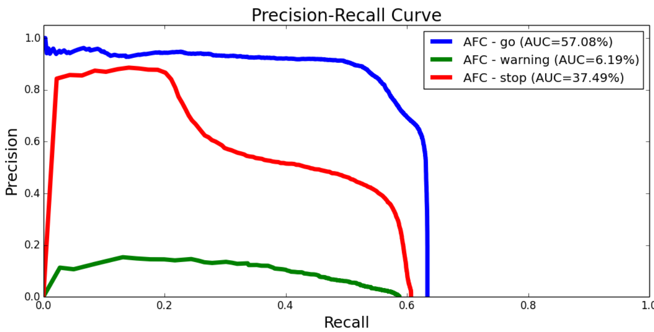


Fig. 10: Recognition performance on skewed dataset.

TABLE VII: Overview of existing public TL databases. The ambiguous class covers uncertain annotations.

	LaRA, Mines ParisTech[70]	*LISA, UCSD*
#Classes	4 (green, orange, red, & ambiguous)	7 (go, go forward, go left, warning, warning left, stop, & stop left)
#Frames/#GT	11,179 / 9,168	43,007 / 119,231
Frame spec.	Mono, 640 x 480, 8-bit, RGB	Stereo, 1280 x 960, 8-bit, RGB
Video	Yes, 8min 49s @ 25FPS	Yes, 44min 41s @ 16FPS
Description	1 sequence, urban, day, Paris, France	4 test seq. $\geq$ 4min and 18 training clips $\leq$ 2min 49s, urban, morning, evening, night, San Diego, USA

The LISA Traffic Light Dataset consists of TLs that are found in San Diego, California, USA. The dataset provides two day and two nighttime sequences for testing. These test sequences contain 23 minutes and 25 seconds of driving around San Diego. The stereo image pairs are acquired using the Point Grey's Bumblebee XB3 (BBX3-13S2C-60) which is constructed with three lenses which each capture images with a resolution of 1280x960. The lenses have a Field of View(FoV) of 66°. Because of the 3 lenses, the stereo camera supports two different baselines, 12 and 24 cm, whereof the widest is used for the LISA Traffic Light Dataset. The stereo images are uncompressed and was rectified on the fly. The Bumblebee XB3 was mounted in the center of the roof of the capturing vehicle and connected to a laptop by FireWire-800 (IEEE-1394b). Besides the 4 test sequences, 18 shorter video clips are provided for training and testing. Gain and shutter speed were manually set to avoid over saturation as well as limit the effect of flickering from the TLs. For all day clips, shutter speed was 1/5000 sec and gain was set to 0. For all night clips, shutter speed was 1/100 sec and gain was set to 8. A Triclops calibration file is provided along with the stereo images, this file contains the factory calibration for the used Bumblebee XB3 camera. Table VIII shows a detailed list of the short video clips and longer video sequences that constitute the LISA Traffic Light Dataset.

The LISA Traffic Light Dataset is captured in stereo since stereo vision is widely used in related computer vision areas and might see more use in TLR. Both mono and stereo vision are widely used for vehicle detection according to [71], which review vehicle detectors. Additionally, [71] describe a 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 [72]. All parts of this paradigm can potentially reduce the amount of false positives. The benefit of stereo vision is reinforced by [73] where the main technical challenges in urban environments are said to be occlusions, shadow silhouettes, and dense traffic. The introduction of stereo has shown promising results in relation to solving these challenges.

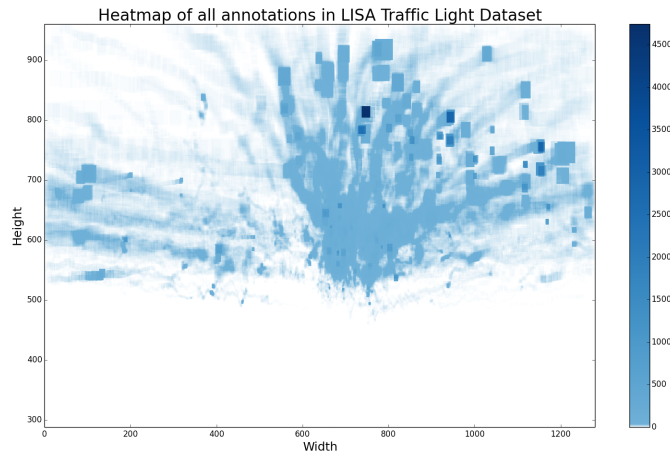


Fig. 11: Heatmap of all annotations in the LISA TL Dataset.

Each sequence in the dataset comes with hand labeled annotations for the left stereo frame. Annotations for a given video sequence contains the following information: frame number, rectangular area around the lit TL lamp, and it's state. A heatmap of the all annotations in the dataset can be seen in Figure 11, where it is clear that most of the annotations are done in the upper right part of the frame, and a few TLs are annotated in the far left side. It is therefore safe to reduce the search for TL to the upper part of the frames.

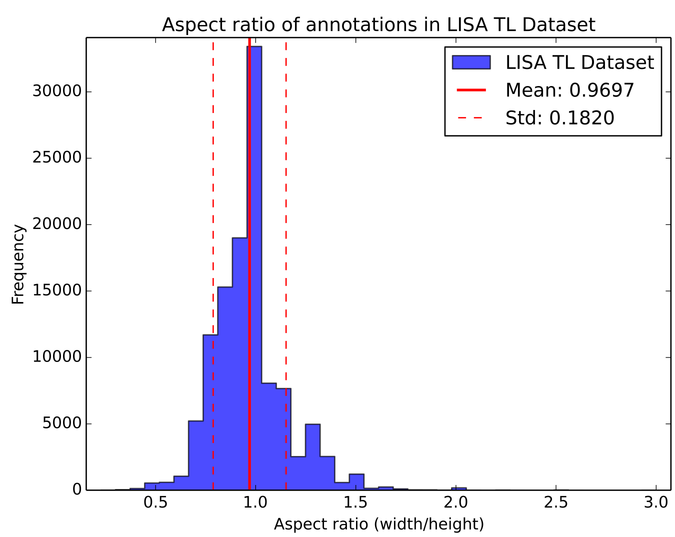


Fig. 12: Aspect ratio histogram of LISA TL Dataset.

Figure 12 shows a histogram of the aspect ratio of all the annotations in the dataset. The mean aspect ratio is 0.9697, which fits the quadratic TL bulbs well. The variation in aspect

TABLE VIII: Overview of the video sequences in LISA Traffic Light Dataset.

Sequence name	Description	# Frames	# Annotations	# TLs	Length	Classes
Day seq. 1	morning, urban, backlight	4,060	10,308	25	4min 14s	Go, warning, warning left, stop, stop left
Day seq. 2	evening, urban	6,894	11,144	35	7min 11s	Go, go forward, go left, warning, stop, stop left
Night seq. 1	night, urban	4,993	18,984	25	5min 12s	Go, go left, warning, stop, stop left
Night seq. 2	night, urban	6,534	23,734	62	6min 48s	Go, go left, warning, stop, stop left
Day clip 1	evening, urban, lens flare	2,161	10,372	10	2min 15s	Go, warning, stop, stop left
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	1,327	3	40s	Go, warning, stop
Day clip 4	evening, urban	398	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	2,719	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 5s	Go, go left, stop, stop left
Day clip 9	morning, urban	960	1,940	4	1min	Go, go left, warning left, stop, stop left
Day clip 10	morning, urban	40	137	4	3s	Go, stop left
Day clip 11	morning, urban	1,053	1,268	6	1min 5s	Go, stop
Day clip 12	morning, urban	152	229	3	9s	Go
Day clip 13	evening, urban	693	1,256	8	43s	Go, warning, stop
Night clip 1	night, urban	591	1,885	8	36s	Go
Night clip 2	night, urban	2,300	4,607	25	2min 23s	Go, go left, warning, warning left, stop, stop left
Night clip 3	night, urban	1,051	2,027	14	1min 5s	Go, go left, warning left, stop, stop left
Night clip 4	night, urban	1,105	2,536	9	1min 9s	Go, warning, stop
Night clip 5	night, urban	1,454	3,242	19	1min 31s	Go, go left, warning, stop, stop left
		43,007	119,231	304	44min 41s	

ratio is caused by viewing angles, motion blur and imprecise annotation.

The LISA Traffic Light Dataset is made freely available at http://cvrr.ucsd.edu/LISA/datasets.html for educational, research, and non-profit purposes.

VIII. Discussion and Perspectives

In this section we discuss the current trends and perspectives based on the surveyed papers. It is difficult to determine the state of TLR as evaluation is done on local datasets and with different evaluation methodology. To maintain and advance research on TLR, it is essential to use a common evaluation methodology on challenging publicly available TL datasets. This will enable both newcomers and established research groups to efficiently compare approaches. [70] provides the only publicly available dataset. It is unfortunately not widely used and lacks variation. The ideal TLR benchmark should have large variation in environmental conditions, similar to *The KITTI Vision Benchmark Suite* for evaluation of stereo correspondence [74], and the *VIVA challenge* [75] for hands, face, and traffic signs. When TLR systems eventually matures, the evaluation metrics should evolve to include weighted penalties for missed or wrong recognitions based on the severity of the error. Furthermore, the distance where TLR systems are first able to successfully recognize a TL is very relevant and should also be part of the evaluation. Since no comprehensive surveys have existed until now, it required substantial effort to gain an overview of the state of TLR research. The scope of existing TLR research vary significantly, spanning from very basic TL detection, to complex systems robust enough to be used in autonomous systems as seen with [32]. Table III indicates that many of the surveyed systems performs in the high 90% in recall, precision, and accuracy. The best performing papers seem to be [40], [32], [35] from the industry, and [54], [52], [58] from academic institutions. The approaches from these papers rely on prior maps of TL location and properties, which makes it possible to achieve solid performance under challenging conditions. Such systems can reduce the number of false positives substantially, because the approximate locations of TLs are known. Using information from precise maps is a big advantage over conventional systems. [58] shows that their use of prior maps increase precision from 56.67% to 97.33%. The price is less flexibility and high cost, since the maps must be kept up to date for the systems to function.

The paradigm change from heuristic models to learning-based seen in traffic sign detection and pedestrian detection has not happened for TLR yet. This is underlined by examining Table I where detection of candidate TLs is almost entirely based on heuristic models, with the exception of two very recent papers. In [58] detection is done using HoG and [59] uses the ACF framework. Learning-based detectors have been tried earlier, but do not appear in Table I, since only the best performing approach from each paper is listed. [34], [40] developed learning-based TL detectors, based on Haar features, to compare with their model-based systems. In both cases a model-based detector outperformed the learning-based detector in both detection and computational load.

The additional information that stereo vision provides is rarely used, one exception to this is [40] where stereo vision is used to measure real world distance and size of detected objects. Doing this resulted in a ten fold decrease in false candidates, along with the tracking benefits of knowing the distance and size. As discussed in [73], stereo vision has proven useful to improve the robustness of computer vision systems. Stereo vision cues could be considered as an additional feature channel or for rejecting false positives from e.g. tail lights and reflections. In [76] vehicle detection at night is assisted by a stereo vision 3D edges extractor, while in [77], vehicle detection rely solely on stereo vision for both day- and nighttime data.

Less than half of the TLR papers include tracking. The most common use of tracking is a simple temporal consistency check. This efficiently suppress FPs and lone FNs. A few papers uses more advanced and sophisticated tracking, such as HMM, IMM, and CAMSHIFT. This is an area that must be researched further as tracking is known to increase performance as seen in [14] where introduction of tracking to traffic signs recognition significantly reduced the number of FPs. For vehicle detection, [78] has similarly increased performance based on vehicle tracking fused with lane localization and tracking.

DAS applications for TLR

There are many applications in which TLR can be used as part of DAS. Table IX lists applications which have been mentioned in the surveyed papers.

Fusion of data from multiple systems and sensors can greatly improve the overall capabilities of DAS. In [27], [26] the driver's attention is measured using cameras looking inside the car. In [25] a first person view camera is used for capturing. The driver's registered attention can e.g. be used to activate safety systems in case of the driver being inattentive. By fusing TL recognition with looking-in systems which e.g detect the driver's eye gaze, it can be determined whether or not the driver have noticed the TL. Other properties which can be used to decide if the driver should be informed are the TL's detectability and discriminability as discussed in [79]. Velocity information from the CAN bus can also be obtained to help determining if the vehicle is slowing down while approaching the TL. A lot of applications require fusion of information from multiple systems, this includes most of the application seen in Table IX. Understanding the traffic scene is necessary as seen with the use of intersection and lane information. A major challenge for TLR in complex intersections is to determine which TLs are relevant to the driver. Selecting the biggest and closest one, as in [80], is a simplistic way of determining which lights to adhere to. In complex intersections, this will not be sufficient and more intelligent approaches must be applied. So far the most intelligent systems for solving this problem is seen in [35], where a guess is made based on the intersection width and the estimated orientation of the TLs. An alternative and less dynamic approach is seen in [32], where relevant TLs are manually annotated before hand. Features are extracted in the annotated regions and the system then recognize relevant TLs on that specific route.

TLR can potentially help people by decreasing fatigue and stress level when driving. This is especially true for people with color vision deficiency or similar challenges. As mentioned in the introduction, a large portion of accidents are connected to intersections and red light running. Integration of TLR systems in vehicles can reduce these accidents. Furthermore, the integration of TLR systems and DAS in cars can to some degree be implemented on smartphones as seen with [81], [41]. Another application for a developed TLR system could be naturalistic driving studies (NDS) analysis by automatic detection of events related to e.g. red light running at intersections. Something similar was done with lane detection in [82], where a set of NDS events are identified and quantified.

Directions

Even though the Daimler group in Germany and the VisLab group in Italy, have successfully managed to make autonomous vehicles drive on public roads, the TLR problem is not considered solved. TLR systems remain challenged by changing weather and light conditions. To overcome these challenges, TLR systems should be able to adapt parameters throughout the TLR pipeline to the changing conditions. Another major problem that still remains to be solved is determining the relevance of recognized TLs. More research should be made into extending TLR for DAS with lane detection, detailed maps, and other traffic scene information. A few learning-based TL detectors have recently been published [58], [59]. It is not possible at this time to tell whether learning-based detectors are superior to heuristic model-based detector. To determine this, more research in applying learning-based detectors for TLR is needed, as well as evaluations on common datasets.

IX. Concluding Remarks

This survey presented an overview of the current state of traffic light recognition (TLR) research in relation to driver assistance systems. The approaches from the surveyed paper were broken down into choices made for color space, detection, features, classification, and tracking. For color space, there exist no clear tendency towards one in particular. We have seen a raising popularity for combining channels from multiple color spaces to create a combined color space that separates traffic light (TL) colors well. Most detection approaches rely on color or shape for finding TL candidates other rely on spotlight detection in a single intensity channel. BLOB analysis is generally used to remove bad TL candidates, this is done based on prior knowledge of the properties of TL BLOBs. Furthermore, some of the best performing approaches use detailed maps of the route and temporal information to improve performance. Many papers utilize manually specified models of TLs, which consist of color, shape, and structural features, to do state detection of TL candidates. Other use trained features such as HoG, LBP, and 2D Gabor wavelets, classified using SVM. A few rely on template matching or neural networks using the color and/or shape. The tracking stage is dominated by temporal filtering, while more advanced approaches include HMM, IMM, and CAMSHIFT.

TLR is dominated by model based approaches, especially for finding TL candidates. This raises the question of whether model based approaches outperform learning based approaches for TLR. Based on the limited experiences with learning based detection this question cannot yet be answered. Additionally, because the systems are evaluated using different methodology and on very different datasets it is not clear

TABLE IX: DAS applications for TLR.

DAS feature	Description	Requirements	References
Dashboard visualization	TL state visualization in dashboard	TLR recognition, lane understanding	[48], [37]
Get going alert	Draw attention to the recently switched light	TLR recognition	[48]
Warn driver of stop light	Drawing attention to upcoming stop light	TLR recognition, intersection, lane understanding	[63], [52], [36]
Stop at stop light	Autonomous vehicle stopping at stop light	TLR recognition, lane understanding	[48]
Smooth stop at stop light	Smooth braking towards stop line at stop light	TLR recognition, stop line, lane understanding	[37]
Stop and go	Automatic stop/start of engine at stop lights	TLR recognition	[48], [37]

which approaches are the best. Only one public dataset with TLs is currently available and it is not widely used. We have therefore contributed a new dataset, the LISA Traffic Light Dataset, which contains TLs captured with a stereo camera in San Diego, USA under varying conditions. The dataset is supposed to enable comparable evaluation on a large and varied dataset, and provides the possibility of including stereo vision for improving TLR. The dataset will be included in the next *VIVA Challenge* [75].

References

[1] J. Sussman, *Perspectives on Intelligent Transportation Systems (ITS)*. Springer US, 2005.

[2] P. Papadimitratos, A. La Fortelle, K. Evenssen, R. Brignolo, and S. Cosenza, "Vehicular communication systems: Enabling technologies, applications, and future outlook on intelligent transportation," *IEEE Communications Magazine*, vol. 47, pp. 84–95, 2009.

[3] Federal Highway Administration. (2009) Traffic control devices: Uses and misuses. [Online]. Available: http://safety.fhwa.dot.gov/intersection/resources/fhwasa10005/brief_3.cfm

[4] AAA Foundation for Traffic Safety. (2014) 2014 traffic safety culture index. [Online]. Available: https://www.aaafoundation.org/sites/default/files/2014TSCIreport.pdf

[5] Federal Highway Administration. (2009) Engineering countermeasures to reduce red-light running. [Online]. Available: http://safety.fhwa.dot.gov/intersection/resources/fhwasa09027/resources/intersection\%20safety\%20issue\%20brief\%206.pdf

[6] D. Shinar, *Traffic Safety and Human Behavior*, ser. Traffic Safety and Human Behavior. Emerald, 2007, no. vb. 5620.

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

[8] The Insurance Institute for Highway Safety (IIHS). (2015) Red light running. [Online]. Available: http://www.iihs.org/iihs/topics/t/red-light-running/topicoverview

[9] Vermont Agency of Transportation. (2009) An evaluation of dilemma zone protection practices for signalized intersection control. [Online]. Available: http://vtransplanning.vermont.gov/sites/aot_program_development/files/documents/materialsandresearch/completedprojects/Dilemma_Zone_Final_Report_6_19_09.pdf

[10] M. Diaz, P. Cerri, G. Pirlo, M. Ferrer, and D. Impedovo, "A survey on traffic light detection," in *New Trends in Image Analysis and Processing – ICIAP 2015 Workshops*, ser. Lecture Notes in Computer Science, V. Murino, E. Puppo, D. Sona, M. Cristani, and C. Sansone, Eds. Springer International Publishing, 2015, vol. 9281, pp. 201–208. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-23222-5_25

[11] M. P. Philipsen, M. B. Jensen, A. Møgelmose, T. B. Moeslund, and M. M. Trivedi, "Ongoing work on traffic lights: Detection and evaluation," *12th IEEE Advanced Video and Signal-based Survaeillance Conference*, 2015.

[12] H. Fleyeh and M. Dougherty, "Road and traffic sign detection and recognition," in *Proceedings of the 16th Mini-EURO Conference and 10th Meeting of EWGT*, 2005, pp. 644–653.

[13] R. O'Malley, M. Glavin, and E. Jones, "Vehicle detection at night based on tail-light detection," in *1st international symposium on vehicular computing systems, Trinity College Dublin*, 2008.

[14] A. Mogelmose, D. Liu, and M. M. Trivedi, "Traffic sign detection for us roads: Remaining challenges and a case for tracking," in *IEEE Transactions on Intelligent Transportation Systems*, 2014, pp. 1394–1399.

[15] M. Mathias, R. Timofte, R. Benenson, and L. Van Gool, "Traffic sign recognition - how far are we from the solution?" in *ICJNN*, 2013.

[16] Y.-L. Chen, C.-T. Lin, C.-J. Fan, C.-M. Hsieh, and B.-F. Wu, "Vision-based nighttime vehicle detection and range estimation for driver assistance," in *IEEE International Conference on Systems, Man and Cybernetics*, 2008, pp. 2988–2993.

[17] C. Idler, R. Schweiger, D. Paulus, M. Mahlisch, and W. Ritter, "Realtime vision based multi-target-tracking with particle filters in automotive applications," in *IEEE Intelligent Vehicles Symposium*, 2006, pp. 188–193.

[18] S. Gormer, D. Muller, S. Hold, M. Meuter, and A. Kummert, "Vehicle recognition and ttc estimation at night based on spotlight pairing," in *12th International IEEE Conference on Intelligent Transportation Systems*, 2009, pp. 1–6.

[19] R. O'Malley, E. Jones, and M. Glavin, "Rear-lamp vehicle detection and tracking in low-exposure color video for night conditions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, pp. 453–462, 2010.

[20] J. C. Rubio, J. Serrat, A. M. López, and D. Ponsa, "Multiple-target tracking for intelligent headlights control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, pp. 594–605, 2012.

[21] R. O'malley, M. Glavin, and E. Jones, "Vision-based detection and tracking of vehicles to the rear with perspective correction in low-light conditions," *IET Intelligent Transport Systems*, vol. 5, pp. 1–10, 2011.

[22] S. Eum and H. G. Jung, "Enhancing light blob detection for intelligent headlight control using lane detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, pp. 1003–1011, 2013.

[23] R. Satzoda and M. Trivedi, "On enhancing lane estimation using contextual cues," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, pp. 1–1, 2015.

[24] A. Bar Hillel, R. Lerner, D. Levi, and G. Raz, "Recent progress in road and lane detection: a survey," *Machine Vision and Applications*, pp. 1–19, 2012.

[25] A. Tawari, A. Mogelmose, S. Martin, T. B. Moeslund, and M. M. Trivedi, "Attention estimation by simultaneous analysis of viewer and view," in *IEEE 17th International Conference on Intelligent Transportation Systems*, 2014, pp. 1381–1387.

[26] A. Tawari, K. H. Chen, and M. M. Trivedi, "Where is the driver looking: Analysis of head, eye and iris for robust gaze zone estimation," in *IEEE 17th International Conference on Intelligent Transportation Systems*, 2014, pp. 988–994.

[27] A. Tawari, S. Sivaraman, M. Trivedi, T. Shannon, and M. Tippelhofer, "Looking-in and looking-out vision for urban intelligent assistance: Estimation of driver attentive state and dynamic surround for safe merging and braking," in *IEEE Intelligent Vehicles Symposium Proceedings*, 2014, pp. 115–120.

[28] California Department of Transportation. (2015) California manual on uniform traffic control devices. [Online]. Available: http://www.dot.ca.gov/hq/traffops/engineering/control-devices/trafficmanual-current.htm

[29] Federal Highway Administration. (2015) Manual on uniform traffic control devices. [Online]. Available: http://mutcd.fhwa.dot.gov/

[30] United Nations. (2006) Vienna convention on road signs and signals. [Online]. Available: www.unece.org/trans/conventn/signalse.pdf

[31] 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, pp. 6919–6934, 2011.

[32] U. Franke, D. Pfeiffer, C. Rabe, C. Knoeppel, M. Enzweiler, F. Stein, and R. Herrtwich, "Making bertha see," in *IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2013, pp. 214–221.

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

[34] R. Charette and F. Nashashibi, "Traffic light recognition using image processing compared to learning processes," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009, pp. 333–338.

[35] N. Fairfield and C. Urmson, "Traffic light mapping and detection," in *Proceedings of ICRA 2011*, 2011.

[36] 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*, pp. 571–574, 2013.

[37] 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*, pp. 3911–3923, 2014.

[38] M. M. Trivedi, T. Gandhi, and J. McCall, "Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety," *IEEE Transactions on Intelligent Transportation Systems*, pp. 108–120, 2007.

[39] E. Ohn-Bar, A. Tawari, S. Martin, and M. M. Trivedi, "On surveillance for safety critical events: In-vehicle video networks for predictive driver assistance systems," *Computer Vision and Image Understanding*, vol. 134, pp. 130 – 140, 2015.

[40] F. Lindner, U. Kressel, and S. Kaelberer, "Robust recognition of traffic signals," in *IEEE Intelligent Vehicles Symposium*, 2004, pp. 49–53.

[41] E. Koukoumidis, M. Martonosi, and L.-S. Peh, "Leveraging smartphone cameras for collaborative road advisories," *IEEE Transactions on Mobile Computing*, vol. 11, pp. 707–723, 2012.

[42] H.-K. Kim, J. H. Park, and H.-Y. Jung, "Effective traffic lights recognition method for real time driving assistance system in the daytime," *World Academy of Science, Engineering and Technology 59th*, 2011.

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

[44] V. Haltakov, J. Mayr, C. Unger, and S. Ilic, "Semantic segmentation based traffic light detection at day and at night," in *Pattern Recognition*, ser. Lecture Notes in Computer Science, J. Gall, P. Gehler, and B. Leibe, Eds. Springer International Publishing, 2015, vol. 9358, pp. 446–457. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-24947-6_37

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

[46] M. Diaz-Cabrera and P. Cerri, "Traffic light recognition during the night based on fuzzy logic clustering," in *Computer Aided Systems Theory-EUROCAST 2013*. Springer Berlin Heidelberg, 2013, pp. 93–100.

[47] M. Diaz-Cabrera, P. Cerri, and J. Sanchez-Medina, "Suspended traffic lights detection and distance estimation using color features," in *15th International IEEE Conference on Intelligent Transportation Systems*, 2012, pp. 1315–1320.

[48] D. Nienhuser, M. Drescher, and J. Zollner, "Visual state estimation of traffic lights using hidden markov models," in *13th International IEEE Conference on Intelligent Transportation Systems*, 2010, pp. 1705–1710.

[49] 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 *17th International Conference on Information Fusion (FUSION)*. IEEE, 2014, pp. 1–7.

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

[51] A. Gomez, F. Alencar, P. Prado, F. Osorio, and D. Wolf, "Traffic lights detection and state estimation using hidden markov models," in *IEEE Intelligent Vehicles Symposium Proceedings*, 2014, pp. 750–755.

[52] Y. Zhang, J. Xue, G. Zhang, Y. Zhang, and N. Zheng, "A multi-feature fusion based traffic light recognition algorithm for intelligent vehicles," in *33rd Chinese Control Conference (CCC)*, 2014, pp. 4924–4929.

[53] C. Jang, C. Kim, D. Kim, M. Lee, and M. Sunwoo, "Multiple exposure images based traffic light recognition," in *IEEE Intelligent Vehicles Symposium Proceedings*, 2014, pp. 1313–1318.

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

[55] 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 *IEEE Intelligent Vehicles Symposium*, 2010, pp. 431–435.

[56] Y. Shen, U. Ozguner, K. Redmill, and J. Liu, "A robust video based traffic light detection algorithm for intelligent vehicles," in *IEEE Intelligent Vehicles Symposium*, 2009, pp. 521–526.

[57] C. Wang, T. Jin, M. Yang, and B. Wang, "Robust and real-time traffic lights recognition in complex urban environments," *International Journal of Computational Intelligence Systems*, vol. 4, no. 6, pp. 1383–1390, 2011.

[58] D. Barnes, W. Maddern, and I. Posner, "Exploiting 3D Semantic Scene Priors for Online Traffic Light Interpretation," in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, Seoul, South Korea, June 2015.

[59] M. P. Philipsen, M. B. Jensen, A. Møgelmose, T. B. Moeslund, and M. M. Trivedi, "Traffic light detection: A learning algorithm and evaluations on challenging dataset," *18th IEEE Intelligent Transportation Systems Conference*, 2015.

[60] M. B. Jensen, M. P. Philipsen, A. Møgelmose, T. B. Moeslund, and M. M. Trivedi, "Traffic light detection at night: Comparison of a learning-based detector and three model-based detectors," *11th Symposium on Visual Computing*, 2015.

[61] M. Omachi and S. Omachi, "Detection of traffic light using structural information," in *IEEE 10th International Conference on Signal Processing (ICSP)*, 2010, pp. 809–812.

[62] G. Siogkas, E. Skodras, and E. Dermatas, "Traffic lights detection in adverse conditions using color, symmetry and spatiotemporal information." in *VISAPP (1)*, 2012, pp. 620–627.

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

[64] S. Naz, H. Majeed, and H. Irshad, "Image segmentation using fuzzy clustering: A survey," in *Emerging Technologies (ICET), 2010 6th International Conference on*, 2010, pp. 181–186.

[65] H. A. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with markovian switching coefficients," *IEEE Transactions on Automatic Control*, vol. 33, pp. 780–783, 1988.

[66] D. Olson and D. Delen, *Advanced Data Mining Techniques*. Springer Berlin Heidelberg, 2008. [Online]. Available: https://books.google.dk/books?id=2vb-LZEn8uUC

[67] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, pp. 303–338, 2010.

[68] C. G. Weng and J. Poon, "A new evaluation measure for imbalanced datasets," in *The 7th Australasian Data Mining Conference-Volume 87*. Australian Computer Society, Inc., 2008, pp. 27–32.

[69] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," in *The 23rd international conference on Machine learning*. ACM, 2006, pp. 233–240.

[70] Robotics Centre of Mines ParisTech. (2015) Traffic lights recognition (tlr) public benchmarks. [Online]. Available: http://www.lara.prd.fr/benchmarks/trafficlightsrecognition

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

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

[73] N. Buch, S. Velastin, and J. Orwell, "A review of computer vision techniques for the analysis of urban traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, pp. 920–939, 2011.

[74] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[75] U. Laboratory for Intelligent and Safe Automobiles . (2015) Vision for intelligent vehicles and applications (viva) challenge. [Online]. Available: http://cvrr.ucsd.edu/vivachallenge/

[76] I. Cabani, G. Toulminet, and A. Bensrhair, "Color-based detection of vehicle lights," in *IEEE Intelligent Vehicles Symposium*, 2005, pp. 278–283.

[77] M. P. Philipsen, M. B. Jensen, R. K. Satzoda, M. M. Trivedi, A. Møgelmose, and T. B. Moeslund, "Night-time drive analysis using stereo-vision for data reduction in naturalistic driving studies," in *IEEE Intelligent Vehicle Symposium*, 2015.

[78] S. Sivaraman and M. M. Trivedi, "Integrated lane and vehicle detection, localization, and tracking: A synergistic approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, pp. 906–917, 2013.

[79] F. Kimura, Y. Mekada, T. Takahashi, I. Ide, H. Murase, and Y. Tamatsu, "Method to quantify the visibility of traffic signals for driver assistance," *IEEJ Transactions on Electronics, Information and Systems*, vol. 130, pp. 1034–1041, 2010.

[80] Y. Jie, C. Xiaomin, G. Pengfei, and X. Zhonglong, "A new traffic light detection and recognition algorithm for electronic travel aid," in *Fourth International Conference on Intelligent Control and Information Processing (ICICIP)*, 2013, pp. 644–648.

[81] E. Koukoumidis, L.-S. Peh, and M. R. Martonosi, "Signalguru: leveraging mobile phones for collaborative traffic signal schedule advisory," in *Proceedings of the 9th international conference on Mobile systems, applications, and services*. ACM, 2011, pp. 127–140.

[82] R. Satzoda and M. Trivedi, "Drive analysis using vehicle dynamics and vision-based lane semantics," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, pp. 9–18, 2015.

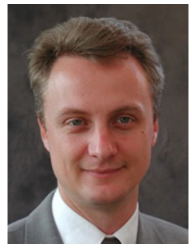
Morten Bornø Jensen received his B.Sc. in Computer Engineering and his M.Sc. in Vision, Graphics, and Interactive Systems from Aalborg University, Aalborg, Denmark, in 2013 and 2015, respectively. He is currently working as Research Assistant at the Visual Analysis of People Laboratory, Aalborg University.

He has been a Visiting Scholar with the Computer Vision and Robotics Research Laboratory, University of California, San Diego. His research interests are computer vision and machine learning.

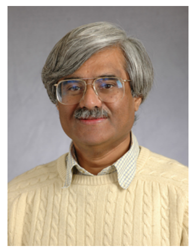
Mark Philip Philipsen received his B.Sc. in Computer Engineering and his M.Sc. in Vision, Graphics, and Interactive Systems from Aalborg University, Aalborg, Denmark, in 2013 and 2015, respectively. He is currently working as Research Assistant at the Visual Analysis of People Laboratory, Aalborg University.

He has been a Visiting Scholar with the Computer Vision and Robotics Research Laboratory, University of California, San Diego. His research interests are computer vision and machine learning.

Andreas Møgelmose received his B.Sc. in Computer Engineering in 2010, his M.Sc. in Informatics in 2012, and his PhD in computer vision for traffic in 2015, all from Aalborg University, Denmark. He is currently a post-doc in the field of automated traffic analysis, also at Aalborg University, where he has worked with everything from traffic sign detection to analysis of radar tracks in shipping lanes. Furthermore, he is a part of the LISA lab at the University of California, San Diego in the Computer Vision and Robotics Research Laboratory, where he has been a visiting scholar multiple times. His main interests are computer vision and machine learning, especially in the area of detecting people, pedestrians, and traffic signs.



Thomas Baltzer Moeslund received the M.Sc.E.E. and Ph.D. degrees from Aalborg University, Aalborg, Denmark, in 1996 and 2003, respectively. He is currently a Professor and the Head of the Visual Analysis of People Laboratory with Aalborg University. He has been involved in ten national and international research projects, as a Coordinator, Work Package leader, and Researcher. He is the author of about 100 peer-reviewed papers. His research interests include all aspects of computer vision, with a special focus on automatic analysis of people. Prof. Moeslund has been a Co-chair of four international workshops/tutorials and a member of the Program Committee for a number of conferences and workshops. He serves as an Associate Editor and as a member of the Editorial Board for four international journals. He received the Most Cited Paper Award in 2009, the Best IEEE Paper Award in 2010, and the Teacher of the Year Award in 2010.



Mohan Manubhai Trivedi received the B.E. (with honors) degree in electronics from Birla Institute of Technology and Science, Pilani, India, in 1974 and the M.S. and Ph.D. degrees in electrical engineering from Utah State University, Logan, UT, USA, in 1976 and 1979, respectively. He is currently a Professor of Electrical and Computer Engineering and the Founding Director of the Computer Vision and Robotics Research Laboratory and the Laboratory for Intelligent and Safe Automobiles (LISA) at the University of California, San Diego. He and his team are currently pursuing research in machine and human perception, multimodal interfaces and interactivity, machine learning, intelligent vehicles, driver assistance and active safety systems. Prof. Trivedi is serving on the Board of Governors of the IEEE Intelligent Society and on the Editorial Board of the IEEE Transportation Systems Transactions on ITS. Prof. Trivedi is a Fellow of the IEEE (for contributions to Intelligent Transportation Systems), Fellow of the IAPR (for contributions to vision systems for situational awareness and human centered vehicle safety), and Fellow of the SPIE (for distinguished contributions to the field of optical engineering).