Aalborg Universitet

Traffic Light Detection

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

Traffic Light Detection: A Learning Algorithm and Evaluations on Challenging Dataset

Mark P. Philipsen[1,2], Morten B. Jensen[1,2],
Andreas Møgelmose[1], Thomas B. Moeslund[1], and Mohan M. Trivedi[2]

Abstract— Traffic light recognition (TLR) is an integral part of any intelligent vehicle, which must function in the existing infrastructure. Pedestrian and sign detection have recently seen great improvements due to the introduction of learning based detectors using integral channel features. A similar push have not yet been seen for the detection sub-problem of TLR, where detection is dominated by methods based on heuristic models.

Evaluation of existing systems is currently limited primarily to small local datasets. In order to provide a common basis for comparing future TLR research an extensive public database is collected based on footage from US roads. The database consists of both test and training data, totaling 46,418 frames and 112,971 annotated traffic lights, captured in continuous sequences under a varying light and weather conditions.

The learning based detector achieves an AUC of 0.4 and 0.32 for day sequence 1 and 2, respectively, which is more than an order of magnitude better than the two heuristic model-based detectors.

I. INTRODUCTION

Recognition of traffic lights (TLs) is an integral part of Driver Assistance Systems (DAS) in the transitional period between manually controlled cars and a fully autonomous network of cars. Currently the focus of research in computer vision systems for vehicles is divided in two. Major industrial research groups, such as Daimlar and Google, are investing heavily in autonomous vehicles and attempt to make computer vision based system for the existing infrastructure. Other research done by academic institutions, such as the LISA lab at UC San Diego and LaRA at ParisTech, are targeting DAS, which is already available to consumers in some high-end models. Existing commercial DAS capabilities include, warning of impending collisions, emergency breaking, automatic lane changing, keeping the advertised speed limit, and adaptive cruise control. For all parts of DAS the urban environment posses a lot of challenges, especially to the systems that rely on computer vision. One of the most important challenge here is detecting and recognizing TLs at intersections. Ideally, the TLs should be able to communicate both visually and using radio communication. However, this requires investments in infrastructure, something that is usually not a high priority.

When some form of computer controlled automation is involved with dangerous objects such as cars, safety and reliability is of utmost importance. The worst case scenarios would be a false positive from e.g. a tail light resulting in the assistance system determining that a red light is imminent when it is not the case and unnecessarily distracting the driver, or worse affecting the driver to perform an emergency braking operation. Most current research is focused on detection and recognition during day time with plenty of light, which makes it much easier to reject false positives, from e.g. tail lights, street lights and various reflections. An exception is a system proposed by Google in [1], where a prior map of the location of TLs makes it possible for their system to achieve solid performance even at night. The same system is able to reduce the number of false positives substantially when it knows where the traffic signal should, and should not be.

Inspiration for further improvements can be found by looking at research done on similar computer vision problems. For sign recognition [2], [3] explain how the focus has shifted from heuristic model-based detection to learning based approaches and the problem is considered solved on a subset of signs. The same is the case with pedestrian detection, where [4] shows how a learning based detectors based on Integral Channel Features (ICF) or the even faster and slightly better Aggregated Channel Features (ACF) outperform the other approaches. While research on sign and pedestrian detection has mostly moved on, the same is not the case for TL detection where the majority rely on some sort of color and/or shape filter for detection.

Research related to pedestrian and traffic signs have benefited greatly from high amount of public datasets made available through various benchmarks, such as the *KITTI Vision Benchmark Suite*[5] and *VIVA Challenge* [6]. Currently only one public TL dataset is available, which is the dataset published by LaRA at ParisTech. The dataset consist of 11,179 frames from a 8min and 49sec long drive in Paris. In order to provide a common basis for comparing future TLR research an extensive public database is collected based on footage from US roads captured under varying light and weather conditions. Each test sequence consists of a continuous drive in an urban environment providing lots of frames with and without TLs.

The purpose of this paper is to compare two heuristic TL detection methods to a state-of-the-art learning based detector relying on ACF. Learning based detectors relying on Haar features have been applied in earlier research [7], [8], [9], without much success. This is therefore the first successful learning based detector applied to the TL detec-

---

[1]Visual Analysis of People Laboratory, Aalborg University, 9000 Aalborg, Denmark.

[2]Computer Vision and Robotics Research Laboratory, UC San Diego, La Jolla, CA 92093-0434, USA

tion problem. Evaluation and comparison between the three approaches is done on daytime sequences from the extensive and difficult LISA Traffic Light Database. The contributions are thus threefold:

1) First successful application of a state-of-the-art learning based detector for TL detection.
2) Comparison between two heuristic TL detection approaches and a learning based detector using ACF.
3) Introduce the first evaluation based on the public LISA Traffic Light Database.

The paper is organized as follows: Relevant research is summarized in section II. In section III we present the proposed methods, followed by evaluation of the TL detectors in section IV. Finally, section V rounds of with some concluding remarks.

II. RELATED WORK

Recent work published in the area of traffic light recognition is reviewed, before developing a traffic recognition system to be used for DAS. For a more extensive overview of the TLR domain, we refer to [10].

A. Traffic Light Recognition

Common for [11], [9], [12] is a TL detector which relies purely on intensity from grayscale images. This has the advantage of being more robust to color distortion. Areas brighter than their surroundings are segmented using the white top-hat mophology operation, which leads to an initial high number of candidates. False candidates are filtered out based on shape information. Specifically, rejection is done based on criteria such as, dimension ratio, the BLOB being free of holes and approximately convex. Furthermore, the areas of BLOBs are compared to the areas of regions grown from extrema in the original grayscale image. This is especially effective for removing false candidates big bright areas such as the sky. This detector relies heavily on a competent classifier for further rejection and state estimation, since the number of false candidates is very high and color information is not available. The detector manages to find 90% of all TLs in the testset.

[13] begins by detecting the vanishing line and thereby reducing the search area considerably, relying on the assumption that TLs will only appear above this line. They then apply the the white top-hat operation as [11], [9], [12] did, on the intensity channel V from a HSV image. What is left is filter based on statistical measurements of the hue and saturation ranges of red and green lights. All pixels outside these ranges are rejected while the remaining pixels are selected as candidates. Remaining BLOBs are filtered based on size and height-width ratio. They then look for black bounding boxes around the BLOBs based on gradient information and the blackness of the inside of box candidates. Their system reaches an accuracy of 85%.

[14] extracts candidate BLOBs from RGB images by applying a color distance transform proposed in [15]. The transform emphasizes the chosen color in an intensity image, which is thresholded to remove to suppressed colors.

This is followed by shape filtering to reduce noise using width/height ratio and the solidity of BLOBs. The solidity is calculated based on the ratio between the area of the BLOB and it's bounding box. When evaluating their system, they count a success if the TL was detected just once in the sequence, this allows them to reach a detection rate of 93,53%.

III. METHODS

In this section all of the methods which are used in the proposed system are presented. The section is divided into two subsection. In the first subsection the learning based detector is described. The second subsection explains the tracking used for improving the output of the detector.

A. Learning based detection

In this subsection we apply the successful ACF detector to the TL detection problem. The learning based detection is similar to the approach seen in [16] for traffic signs. We use the Matlab toolbox provided by [17]. The learning based detection system is described in the following three parts:

1) Features: The learning based detector is based on features from 10 channels as described in [18]. A channel refers to a representation of the input image. The 10 different channels include 6 gradient histogram channels, 1 for unoriented gradient magnitude, and 3 for the channels in the CIE-LUV color space. In each channel, small rectangular blocks are used as features. These features are evaluated using a modified AdaBoost classifier with depth-2 decision trees as weak learners.

2) Training: Training is done using 14,106 positive TL samples with a resolution of 20x40 and 42,125 negative samples from 200 carefully selected frames without TLs. In Figure 1 four examples of the positives used for the learning based detector are seen. Similarly, Figure 2 shows two examples of frames used for negatives.
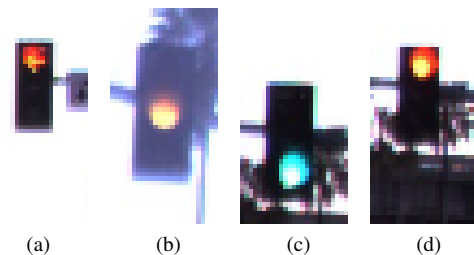

(a)   (b)   (c)   (d)

Fig. 1: Positive samples for learning based detector.

The classifier is trained with Adaboost based on the features extracted from the positive samples. We train 4 cascade stages, 1st stage consists of 10 weak learners, 2nd stages of 100, 3rd stage of 1000, and 4th stage of 20000. In the 4th stage, the training algorithm convergent at 3136 weak learners.

3) Detection: We use a 20x40 sliding window across an integral image of each of the 10 channels in the test image.
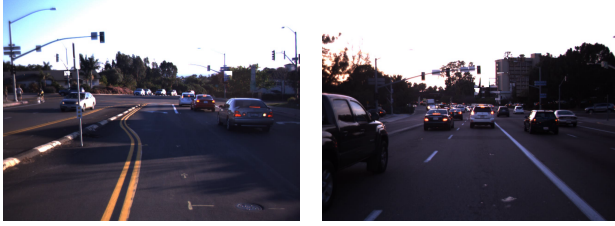
Fig. 2: Negative samples for learning based detector.

B. Heuristic model based detection

We want to compare the learning based detector to more conventional detectors based on heuristic models. The first approach is based on back projection of trained color histograms of the three TL colors. The second approach is purely relying on intensity information for spotlight detection.

1) Detection by Back Projection: Back projection begins with the generation of color distribution histograms. These histograms are created from 10 specifically selected training samples for each color, green, yellow, and red. Based on the U and V channels of the LUV color space a 2D histogram is created for each of the colors. The histograms are min-max normalized before they are used for back projection. The resulting back projection is thresholded to remove low probability pixels. TLs are found using BLOB analysis, and size, shape information is used to generate confidence scores for each BLOB. The specific metrics are listed here:

- Ratio between width and height of bounding box
- Mean value inside bounding box in the back projection image
- Mean value inside bounding box in the intensity image
- Ratio between area of floodfilled BLOB and area of bounding box

2) Detection by Spotlight Detection: Spotlights are found in the intensity channel L from the LUV colorspace using the white top-hat morphology operation. This method has been used in a significant fraction of recent TLR papers [11], [9], [12], [13], [19]. The found spotlight are scored based on the listed metrics.

- Ratio between width and height of bounding box
- Ratio between the convex area of BLOB and area of bounding box
- Ratio between area of floodfilled BLOB and area of bounding box

IV. EVALUATION

The systems are evaluated based upon the following five criteria:

- 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 on Precision-Recall curves

$$Precision = \frac{TP}{TP + FP} \qquad (1)$$

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

$$Recall = \frac{TP}{TP + FN} \qquad (2)$$

Recall is the ratio of correct TL detections compared to the total number of detections.

For presenting and evaluating the overall system performance, we use a precision-recall curve and using the area-under-curve (AUC) as measure. A high AUC indicates good performance, an AUC of 100% indicates a perfect system for the testset.

All systems are evaluated on the two test day sequences from the LISA Traffic Light Database[1]. This provides a total of frame number of 14,386, and a total ground truth of 21,421 annotated TLs. Additional information of the video sequences can be found in Table I. The resolution of the LISA Traffic Light Database is 1280x960. Only the upper 1280x580 part of the frames are used, which results in a system evaluation time of an average 1.275 seconds per frame. We present the results according to the orignial PASCAL overlap criteria of 50 % in Figure 3 and 4.
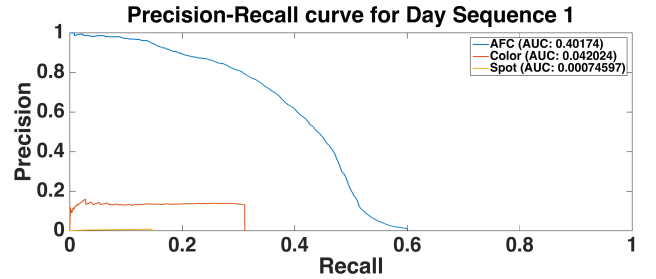

Fig. 3: Precision-Recall curve of day sequence 1 using 50 % overlap criteria.
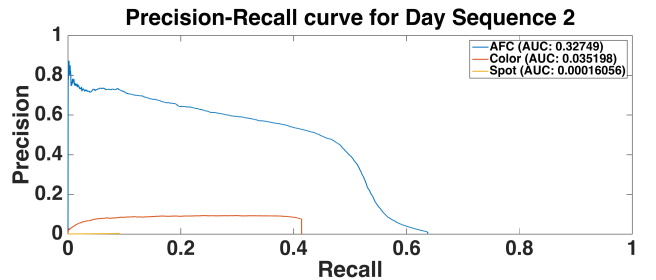

Fig. 4: Precision-Recall curve of day sequence 2 using 50 % overlap criteria.

By examining figure 3 and 4, it is clear that the learning based detector far outperforms the other detectors in both precision and recall when evaluated on both day sequences. During evaluation especially the spotlight detector would

TABLE I: Overview of the daytime test sequences in LISA Traffic Light Database.

Sequence name	Description	# Frames	# Annotations	# TLs	Length
Day sequence 1	morning, urban	4,800	10,267	25	5.00 min
Day sequence 2	evening, urban	9,586	11,154	29	6.10 min
		14,386	21,421	54	11.1 min

miss a lot of otherwise correct detection because of the harsh overlap criteria. The primary reason for this being the inaccuracy in estimating the TL box from the detected spotlights.

To show the impact of these inaccuracies, the system is also validated using a more gentle overlap criterion of 25 %. The results from are presented in Figure 5 and 6.
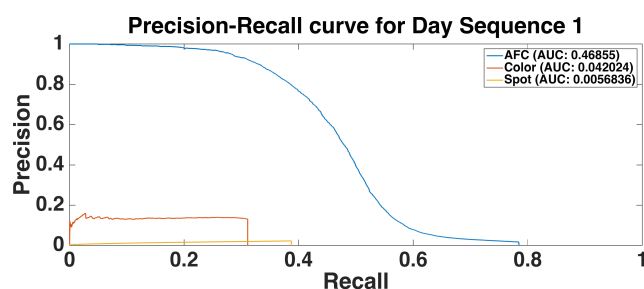

Fig. 5: Precision-Recall curve of day sequence 1 using 25 % overlap criteria.
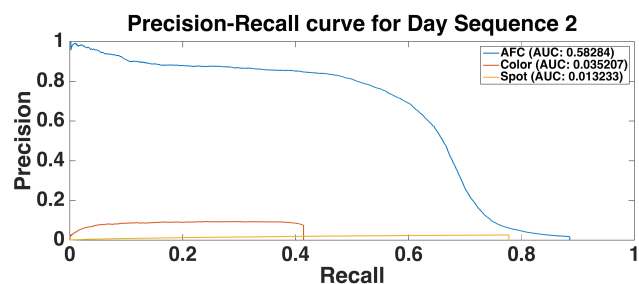

Fig. 6: Precision-Recall curve of day sequence 2 using 25 % overlap criteria.

Easing of the overlap criterion shows significantly improved AUC for all the detectors. It is therefore apparent that improvements in determining location and scale is necessary. From Figures 3, 4, 5, and 6 it seems that the confidence metrics defined in subsection III-B for the model-based detectors are bad at discriminating between TLs and non TL spotlights. It is apparent that for especially the spotlight detector false candidates obtain a better score than actual TLs. The learning based approach is trained towards detecting the entire TL rather than only the TL spot, which gives it an advantage compared to the two model based which are optimized towards the TL spot.

In Figure 7 two detection images from the learning based system is seen. The green bounding box is the positive detected TLs, and the red bounding box is false positives. The true positive detected TLs have a score around 400, and the false positives have a score around 200 making it easy to discard them.

(a)

(b)

Fig. 7: Detections by the learning based detector.

V. CONCLUDING REMARKS

We have compared a learning based detector based on aggregated channel features to two detectors based on heuristic models. The learning based detector reached the best AUC, because of the significantly higher precision and recall. For detectors recall is usually the most important parameter, since many of the false positives can be removed in later stages, whereas false negatives are lost for good. The learning based detector achieves an AUC of 0.4 and 0.32 for day sequence 1 and 2, respectively. This is more than an order of magnitude better than the two heuristic model-based detectors.

On top of the detectors we would like to implement tracking to reduce the number of false positives and false negatives. Stereo vision could be used to filter out false positives by looking at the detected TL candidates' height above the road surface as well as their size and shape. 3D information can also be used to improve tracking precision.

REFERENCES

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

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

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

[4] P. Dollár, R. Appel, S. Belongie, and P. Perona, "Fast feature pyramids for object detection," *PAMI*, 2014.

[5] 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.

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

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

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

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

[10] M. P. Philipsen, M. B. Jensen, M. M. Trivedi, A. Møgelmose, and T. B. Moeslund, "Vision for looking at traffic lights: Issues, survey, and perspectives," in *Intelligent Transportation Systems, IEEE Transactions [In submission]*. IEEE, 2015.

[11] G. Trehard, E. Pollard, B. Bradai, and F. Nashashibi, "Tracking both pose and status of a traffic light via an interacting multiple model filter," in *Information Fusion (FUSION)*. IEEE, 2014, pp. 1–7.

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

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

[14] 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.

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

[16] A. Mogelmose, D. Liu, and M. Trivedi, "Traffic sign detection for u.s. roads: Remaining challenges and a case for tracking," in *Intelligent Transportation Systems (ITSC), IEEE*, 2014, pp. 1394–1399.

[17] P. Dollár, "Piotr's Computer Vision Matlab Toolbox (PMT)," http://vision.ucsd.edu/~pdollar/toolbox/doc/index.html.

[18] P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral channel features." in *BMVC*, vol. 2, 2009, p. 5.

[19] D. Nienhuser, M. Drescher, and J. Zollner, "Visual state estimation of traffic lights using hidden markov models," in *Intelligent Transportation Systems (ITSC), IEEE*. IEEE, 2010, pp. 1705–1710.