

Traffic Light Detection at Night: Comparison of a Learning-Based Detector and Three Model-Based Detectors

Morten B. Jensen^{1,2(✉)}, Mark P. Philipsen^{1,2}, Chris Bahnsen¹,
Andreas Møgelmoose^{1,2}, Thomas B. Moeslund¹, and Mohan M. Trivedi²

¹ Visual Analysis of People Laboratory, Aalborg University, Aalborg, Denmark
mboj@create.aau.dk

² Computer Vision and Robotics Research Laboratory, UC San Diego,
La Jolla, USA

Abstract. Traffic light recognition (TLR) is an integral part of any intelligent vehicle, it must function both at day and at night. However, the majority of TLR research is focused on day-time scenarios. In this paper we will focus on detection of traffic lights at night and evaluate the performance of three detectors based on heuristic models and one learning-based detector. Evaluation is done on night-time data from the public LISA Traffic Light Dataset. The learning-based detector outperforms the model-based detectors in both precision and recall. The learning-based detector achieves an average AUC of 51.4% for the two night test sequences. The heuristic model-based detectors achieves AUCs ranging from 13.5% to 15.0%.

1 Introduction

Traffic lights are used to safely regulate the traffic flow in the current infrastructure, they are therefore a vital part of any intelligent vehicle, whether it is fully autonomous or employ Advanced Driver Assistance Systems (ADAS). In either application, TLR must be able to perform during both day and night. TLR for night-time scenarios is especially important as more than 40% of accidents at intersections occur during the late-night/early-morning hours, in fact a crash is 3 times more probable during the night than during the day [1]. For more introduction to TLR in general we refer to [2] where an overview is given of the current state of TLR. In the same paper, the lack of a large public dataset is addressed with the introduction of the LISA Traffic Light Dataset, which contains challenging conditions and both day- and night-time data.

Before the state of traffic lights (TLs) can be determined they must first be detected. Traffic light detection (TLD) has proven to be very challenging under sub-optimal and changing conditions. The purpose of this paper is therefore to evaluate the night-time TLD performance of three heuristic TL detectors and compare this to a state-of-the-art learning based detector relying on Aggregated Channel Features (ACF). The same learning-based detection framework has previously been applied for day-time TLD in [3]. This makes it possible to compare

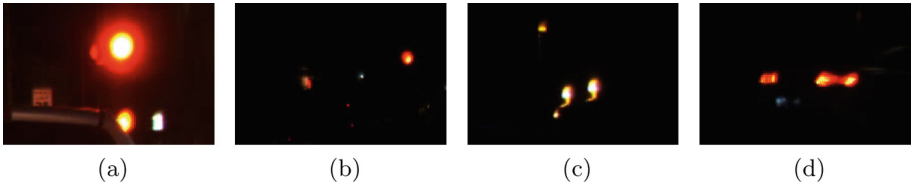


Fig. 1. Challenges samples from the LISA Traffic Light Dataset.

the detector's performance at night and day. Evaluation is done on night-time 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 TLD at night.
2. Comparison of three model-based TLD approaches and a learning-based detector using ACF.
3. Clarification of the challenges for night-time TLD.

The paper is organized as follows: Challenges specific to night-time TLD are clarified in Sect. 2. Relevant research is summarized in Sect. 3. In Sect. 4 we present the detectors, followed by the evaluation in Sect. 5. Finally, Sect. 6 rounds off with our concluding remarks.

2 Traffic Lights and Their Variations

In this section we present some challenges particular to night-time TLD.

1. Lights may seem larger than the actual source [4], see Fig. 1a.
2. Colors saturate to white [4], see Fig. 1a.
3. Lack of legal standards for tail-lights in the USA, tail-lights may therefore resemble TLs [5], see Fig. 1d.
4. TL may be reflected in reflective surfaces, e.g. storefronts, see Fig. 1b.
5. Street lamps and other light sources may look similar to TLs, see Fig. 1c.

Type 1 and 2 can be reduced by increasing the shutter speed at the risk of getting underexposed frames. One solution to this problem is seen in [6], where frames are captured by alternating between slow and fast shutter speed. Generally, it is hard to cope with the remaining issues from a detection point of view. One solution to removing type 3, 4 and 5 false positives could be the introduction of prior maps with information of where TLs are located in relation to the ego-vehicles location, as e.g. seen in [5].

3 Related Work

Most research on TLD and TLR has been focused on day-time, only a handful of publications evaluate their systems on night-time data. One is [4] where a fuzzy clustering approach is used for detection. Gaussian distributions are calculated based on the red, amber, green, and black clusters in a large number of combinations of the RGB and RGB-N image channels. In [7] the work from [4] is expanded, by the introduction of an adaptive shutter and gain system, advanced tracking, distance estimation, and evaluate on a large and varied dataset with both day-time and night-time frames. Because of the differences in light conditions between night and day, they use one fuzzy clustering process for day conditions and another for night conditions. [8] finds TL candidates by applying the color transform proposed in [9]. The color transform determines the dominant color of each pixel based on the RGB values. Dominant color images are only generated for red and green, since no transform is presented for yellow. After thresholding of the dominant color images, BLOBs are filtered based on the width to height ratio and the ratio between the area of the BLOB and the area of the bounding box. The remaining TL candidates are then classified using SVM on a wide range of BLOB features.

When looking at TL detectors which have been applied to day-time data, two recent papers have employed learning-based detectors. [10] 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. [3] uses the ACF framework provided by [11]. Here 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. Spotlight detection using the white top hat operation on intensity images is seen in [12–15]. In [16], the V channel from the HSV color space is used with the same effect. A high proportion of publications use simple thresholding of color channels in some form. [6] is a recent example where traffic light candidates are found by setting fixed thresholds for red and green TL lamps in the HSV color space.

For a more extensive overview of the TLR domain, we refer to [2].

4 Methods

In this section we present the used methods. In the first subsection the learning-based detector is described. The second describes each of the three model-based detectors and how the confidence scores are calculated for the TL candidates found by these model-based detectors.

4.1 Learning-Based Detection

In this subsection we describe how the successful ACF detector has been applied to the night-time TL detection problem. The learning-based detector is provided

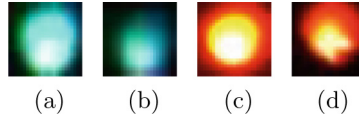


Fig. 2. Positive samples for training the learning-based detector.



Fig. 3. Negative samples for training the learning-based detector.

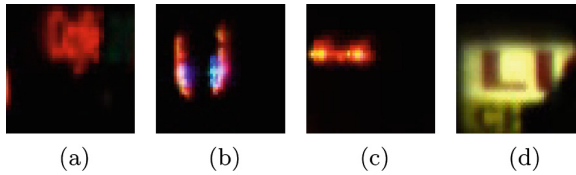


Fig. 4. Hard negative samples for training the learning-based detector.

as part of the Matlab toolbox from [11]. It is similar to the detectors seen in [17] for traffic signs and [3] for day-time TLs, except for few differences in the configuration and training which are described below.

Features. The learning-based detector is based on features from 10 channels as described in [18]. A channel is a representation of the input image, which is obtained by various transformations. The 10 different channels include 6 gradient histogram channels, 1 for unoriented gradient magnitude, and 3 for each channels in the CIE-LUV color space. In each channel, the sums of small blocks are used as features. These features are evaluated using a modified AdaBoost classifier with depth-4 decision trees as weak learners.

Training. Training is done using 7,456 positive TL samples with a resolution of 25×25 and 163,523 negative samples from 5,772 selected frames without TLs. Figure 2 shows four examples of the positive samples used for training the detector. Similarly, Fig. 3 shows two examples of frames used for negative samples. Finally, Fig. 4 shows four hard negative samples extracted using false positives from the training dataset.

AdaBoost is used to train 3 cascade stages, 1st stage consists of 10 weak learners, 2nd stages of 100, and 3rd stage is set to 4,000 but converges at 480.

In order to detect TLs at a greater interval of scales, the octave up parameter is set to 1 instead of the default 0. The octave up parameters defines the number of octaves to compute above the original scale.

Detection. A 18×18 sliding window is used across each of the 10 aggregated channels in the frames from the test sequences.

4.2 Heuristic Model-Based Detection

We want to compare the learning-based detector to more conventional detector types which are based on heuristic models. For each of the three model-based detectors, a short description is given along with output showing central parts of the detectors. The sample shown in Fig. 1a is used as input.

Detection by Thresholding. The detector which uses thresholding is mainly based on the work presented in [6]. Thresholds are found for each TL color in the HSV color space by looking at values of individual pixels from TL bulbs sampled from the training clips in the LISA Traffic Light dataset. Figure 5(a) shows the input sample and Fig. 5(b) shows output after thresholding input. Pixels that fall inside the thresholds for one of the three colors are labeled green, yellow or red in Fig. 5. For the input sample only pixels which fell within the yellow and red thresholds were present.

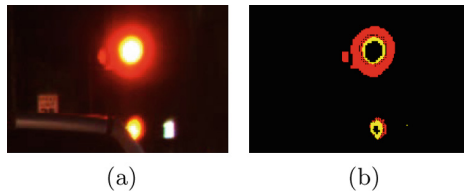


Fig. 5. Thresholded TL (Color figure online).

Detection by Back Projection. Back projection begins with the generation of color distribution histograms. The histograms are two-dimensional and are created for each of the TL colors using 20 training samples for each of the TL colors, green, yellow, and red. From the training samples the U and V channels of the LUV color space are used. The histograms are normalized and used to generate a back projection which is thresholded to remove low probability pixels from the TL candidate image. The implementation is similar to our previous work in [3]. Figure 6a shows the back projected TL candidate image. Figure 6b shows the processed back projected TL candidate image after removal of low probability pixels and some typical morphology operations.

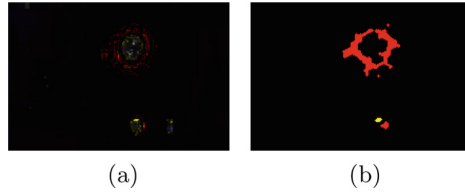


Fig. 6. Back projected TL (Color figure online).

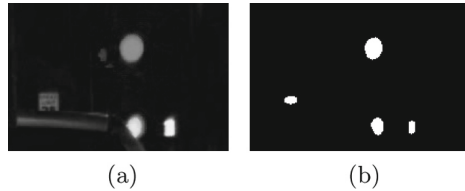


Fig. 7. Spotlights found using the white top-hat operation.

Detection by Spotlight Detection. Spotlights are found in the intensity channel L from the LUV colorspace using the white top-hat morphology operation. The implementation is similar to our previous work in [3]. This method has been used in a many recent TLR papers [12–16]. Figure 7a shows the output of the white top-hat operation. Figure 7b shows the binarized TL candidate image after thresholding and some typical morphology operations.

Confidence Scores for TL Candidates. Confidence scores are calculated for all TL candidates found by the three model-based detectors. The TL BLOB characteristics used in this work have seen use in earlier work, such as [8, 9]. Scores from individual characteristics are generated ranging from [0–1], with 1 being the best. These are summed for each TL candidate, resulting in a combined score ranging from [0–5].

Bounding box ratio: The bulbs of TLs are circular, therefore under ideal conditions the bounding box will be quadratic. The bounding box ratio is calculated as the ratio between width and height of the bounding box.

Solidity ratio: Since TL bulbs are captured as circular and solid under ideal conditions, a BLOBs solidity is a characteristic feature for a TL. The solidity is calculated as the ratio between the convex area of detected BLOBs and the area of a perfect circle, with a radius approximated from the dimensions of the BLOB.

Mean BLOB intensity: Each of the three detectors produce an intensity channel which can be interpreted as a confidence map of TL pixels. The best example is from detection by back projection, where the result of the back projection

Table 1. Overview of night test sequences from the LISA Traffic Light Dataset.

Sequence name	Description	# Frames	# Annotations	# TLs	Length
Night seq. 1	night, urban	4,993	18,984	25	5 min 12 s
Night seq. 2	night, urban	6,534	23,734	62	6 min 48 s
		11,527	42,718	87	12 min

is an intensity channel with normalized probabilities of each pixel being a TL pixel. The intensity channel employed from the spotlight detector is less informative, since it describes the strength of the spotlight. From the threshold based detector, we simply use the intensity channel from the LUV color space.

Flood-filled area ratio: The bulbs of TLs are surrounded by darker regions, by applying flood filling from a seed inside the found BLOBs, it can be confirmed that this contrast exists. We use the ratio between the area of the bounding box and the area of the bounding box from the flood filled area as a measure for this.

Color confidence: Using basic heuristics based thresholding we find the most prominent color inside the TL candidates' bounding boxes. The confidence is calculated based on the number of pixels belonging to that color and the total number of pixels within the bounding box. Pixels with very low saturation are not included in the confidence calculation.

5 Evaluation

Most TL detectors have been evaluated on datasets which are unavailable to the public. This makes it difficult to determine the quality of the published results and compare competing approaches. We strongly advocate that evaluation is done on public datasets such as the LISA Traffic Light Dataset¹.

5.1 LISA Dataset

The four detectors presented in this paper are evaluated on the two night test sequences from the LISA Traffic Light Dataset. This provides a total of 11,527 frames, and a total ground truth of 42,718 annotated TL bulbs. Additional information of the video sequences can be found in Table 1. The resolution of the LISA Traffic Light Database is 1280×960 , however only the upper 1280×580 pixels are used in this paper.

¹ Freely available at <http://cvrr.ucsd.edu/LISA/datasets.html> for educational, research, and non-profit purposes.

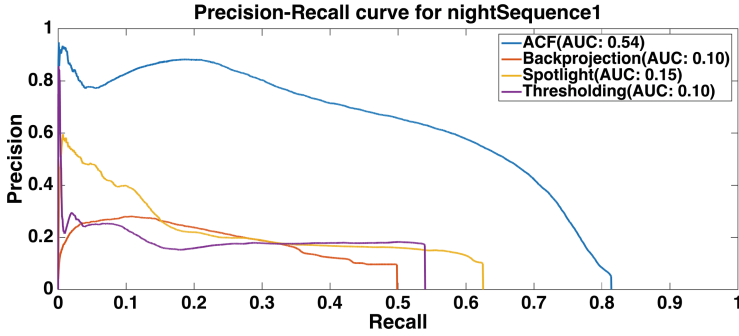


Fig. 8. Precision-Recall curve of night sequence 1 using 50% overlap criteria.

5.2 Evaluation Criteria

In order to insure that the evaluation of TL detectors provide a comprehensive insight into the detectors performance, it is important to use descriptive and comparable evaluation criteria. The presented detectors are evaluated based upon the following four criteria:

PASCAL overlap criterion defines a true positive (TP) to be a detection with more than 50% overlap over ground truth (GT).

Precision is defined in Eq. (1).

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

Recall is defined in Eq. (2).

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

Area-under-curve (AUC) for a precision-recall (PR) curve is used as a measure for the overall system performance. A high AUC indicates good performance, an AUC of 100% indicates perfect performance for the testset.

5.3 Results

We present the final results according to the original PASCAL overlap criteria of 50% in Figs. 8 and 9.

By examining Figs. 8 and 9, it is clear that the learning-based detector outperforms the other detectors in both precision and recall on both night sequences. The odd slopes of the PR curves for the back projection detectors are a result of problems with getting filled and representative BLOBs. The learning-based

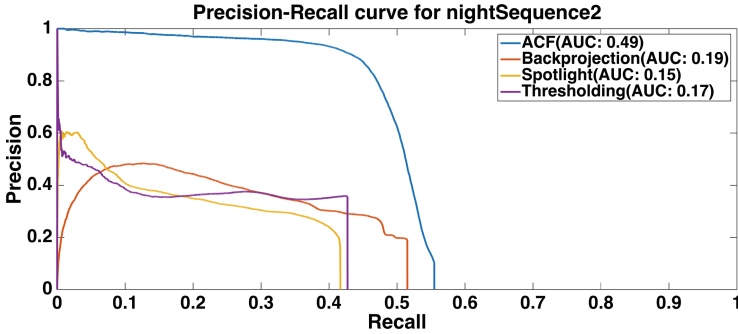


Fig. 9. Precision-Recall curve of night sequence 2 using 50% overlap criteria.

detector is able to differentiate well between TLs and other light sources, leading to a great precision and smooth precision-recall curve. The main problems with the learning-based detector's PR curves are the false negatives caused by detections not meeting the PASCAL criteria but still reaching a very high score, and problems with detecting TLs from far away. These detections causes some instability in the precision especially around 0.05 recall in Fig. 8.

6 Concluding Remarks

We have compared three detectors based on heuristic models to a learning-based detector based on aggregated channel features. The learning-based detector reached the best AUC because of the significantly higher precision and good recall. Recall is generally seen as the most important performance metric for detectors since precision can be improved in later stages, whereas false negatives are lost for good. The learning-based detector achieves an average AUC of 51.4% for the two night test sequences. The heuristic model-based detectors achieved average AUCs ranging from 13.5% to 15.0%, with detection by back projection and spotlight detection achieving the highest AUCs.

Interesting future TLD work could involve applying and comparing the performance of deep learning methods on the LISA TL Dataset with the results presented in this paper.

References

1. Federal Highway Administration: Reducing late-night/early-morning intersection crashes by providing lighting (2009)
2. Jensen, M.B., Philipsen, M.P., Trivedi, M.M., Møgelmoose, A., Moeslund, T.B.: Vision for looking at traffic lights: Issues, survey, and perspectives. *IEEE Trans. Intell. Transport. Syst.* IEEE (2015, in submission)

3. Philipsen, M.P., Jensen, M.B., Møgelmoose, A., Moeslund, T.B., Trivedi, M.M.: Traffic light detection: a learning algorithm and evaluations on challenging dataset. In: 18th International Conference on IEEE Intelligent Transportation Systems Conference, pp. 2341–2345 (2015)
4. Diaz-Cabrera, M., Cerri, P.: Traffic light recognition during the night based on fuzzy logic clustering. In: Moreno-Díaz, R., Pichler, F., Quesada-Arencibia, A. (eds.) EUROCAST. LNCS, vol. 8112, pp. 93–100. Springer, Heidelberg (2013)
5. Fairfield, N., Urmsom, C.: Traffic light mapping and detection. In: Proceedings of ICRA 2011, pp. 5421–5426 (2011)
6. Jang, C., Kim, C., Kim, D., Lee, M., Sunwoo, M.: Multiple exposure images based traffic light recognition. In: IEEE Intelligent Vehicles Symposium Proceedings, pp. 1313–1318 (2014)
7. Diaz-Cabrera, M., Cerri, P., Medici, P.: Robust real-time traffic light detection and distance estimation using a single camera. *Expert Syst. Appl.* **42**(8), 3911–3923 (2015)
8. Kim, H.K., Shin, Y.N., Kuk, S.G., Park, J.H., Jung, H.Y.: Night-time traffic light detection based on SVM with geometric moment features. In: 76th World Academy of Science, Engineering and Technology, pp. 571–574 (2013)
9. Ruta, A., Li, Y., Liu, X.: Real-time traffic sign recognition from video by class-specific discriminative features. *Pattern Recogn.* **43**, 416–430 (2010)
10. Barnes, D., Maddern, W., Posner, I.: Exploiting 3D semantic scene priors for online traffic light interpretation. In: Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Seoul, South Korea (2015)
11. Dollár, P.: Piotr's Computer Vision Matlab Toolbox (PMT) (2015). <http://vision.ucsd.edu/~pdollar/toolbox/doc/index.html>
12. Trehard, G., Pollard, E., Bradai, B., Nashashibi, F.: Tracking both pose and status of a traffic light via an interacting multiple model filter. In: 17th International Conference on Information Fusion (FUSION), pp. 1–7. IEEE (2014)
13. Charette, R., Nashashibi, F.: Traffic light recognition using image processing compared to learning processes. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 333–338 (2009)
14. de Charette, R., Nashashibi, F.: Real time visual traffic lights recognition based on spot light detection and adaptive traffic lights templates, In: IEEE Intelligent Vehicles Symposium, pp. 358–363 (2009)
15. Nienhuser, D., Drescher, M., Zollner, J.: Visual state estimation of traffic lights using hidden markov models. In: 13th International IEEE Conference on Intelligent Transportation Systems, pp. 1705–1710 (2010)
16. Zhang, Y., Xue, J., Zhang, G., Zhang, Y., Zheng, N.: A multi-feature fusion based traffic light recognition algorithm for intelligent vehicles. In: 33rd Chinese Control Conference (CCC), pp. 4924–4929 (2014)
17. Mogelmoose, A., Liu, D., Trivedi, M.M.: Traffic sign detection for us roads: Remaining challenges and a case for tracking. In: Intelligent Transportation Systems, pp. 1394–1399. IEEE (2014)
18. Dollár, P., Tu, Z., Perona, P., Belongie, S.: Integral channel features. In: BMVC, vol. 2, p. 5 (2009)