Robust recognition of traffic signals

3 authors, including:

Frank Lindner
Daimler
12 PUBLICATIONS   **1,086** CITATIONS

Ulrich Kreßel
Daimler
41 PUBLICATIONS   **1,400** CITATIONS

Robust Recognition of Traffic Signals

Frank Lindner, Ulrich Kressel, and Stephan Kaelberer

Abstract—**In this paper a general system for real-time detection and recognition of traffic signals is proposed. The key sensor is a camera installed in a moving vehicle. The software system consists of three main modules: detection, tracking, and sample-based classification. Additional sensor information, such as vehicle data, GPS, and enhanced digital maps, or a second camera for stereo vision, are used to enhance the performance and robustness of the system.**

Since the detection step is the most critical one, different detection schemes are compared. They are based on color, shape, texture and complete-object classification. The color system, with a high dynamic range camera and precise location information of the vehicle and the searched traffic signals, offers valuable and reliable help in directing the driver's attention to traffic signals and thus can reduce red-light running accidents.

I. INTRODUCTION

FOR advanced driver assistance systems the urban scenario is still a challenging, but − by avoiding accidents − rewarding task [1]. A closer look to (German) accident statistics shows, that about two-thirds of all accidents within city boundaries occur at intersections. About 50% of these accidents occur despite traffic signals. The accident severeness is even overproportional as can be seen from the number of injuries and casualties [2].

Previous work [1] demonstrated the feasibility of recognizing traffic lights using a color camera, but limitations of the color sensor and the lack of GPS and enhanced digital maps did not allow to build a robust system at that time. Furthermore, since the first generation of car cameras e.g. for lane assistance or night vision (will) have gray value sensors only, the question has to be asked, whether and how the missing key feature 'color' degrades the performance of a traffic signal recognition system.

The outline of the paper is as follows: Section II introduces the general system architecture proposed for traffic signal recognition. In section III a complete system based on color detection is presented. Section IV discusses several

detection methods for a gray value camera system. The results are summarized in section V leading to an overall system evaluation.

II. SYSTEM OVERVIEW

The basic layout of the traffic signal recognition system consists of three main modules. First, an interest operator (detector) – using either color, shape or/and texture as key features – scans each frame for traffic signal hypotheses. Secondly, a tracker groups these candidates over several frames to form an object/track hypothesis. Finally, a classifier verifies each single candidate, whether it is a traffic signal or not, and also classifies the status of the traffic signal (red, yellow or green).

This three stage system layout already proved its effectiveness in traffic sign recognition [3] and needs a single camera and a processing unit as basic setup. Additional sensor information, such as vehicle data (velocity and yaw rate), GPS and enhanced digital maps, or a second camera for stereo vision, may be used to improve the performance and robustness of the complete system.
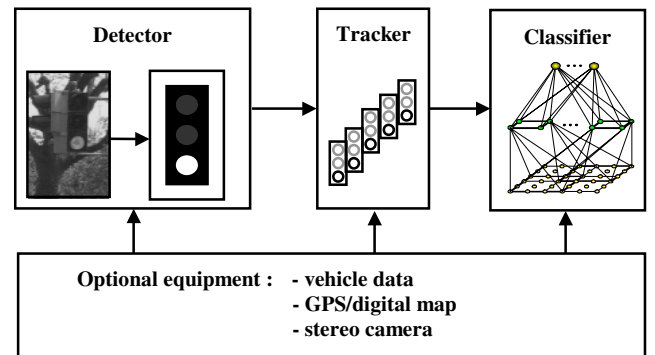

Fig. 1. System architecture for the traffic signal recognition system: detector, tracker, and classifier, supported by other information.

A. Detector

The first stage of the system, and the most critical one, is the detector, since it has to cope with a large amount of input data. For real time applications local operators are preferable, which even might be implemented on FPGA hardware in the car. On the one hand, the detection step should miss only very few traffic lights, since these gaps

have to be filled by the tracking module. On the other hand, quite a few false candidates can be tolerated, which can be suppressed by the tracker or/and the classifier.

For a color camera system a single-pixel classifier, compressed into a fast look-up table, is sufficient, as shown in more detail in section III. Using a gray value camera, the glowing traffic light is the most prominent feature. It can be detected either by its shape (circle detector, see section IV-A) or its texture (blob detector, see section IV-B). Nowadays, even complete-object classifiers are becoming affordable as detectors with respect to computational complexity, which will be discussed in more detail in section IV-C.

B. Tracker

The tracker groups detected candidates from consecutive images. It takes into account the relative location, speed and acceleration between the moving camera and the traffic signal. It compensates not only some drop outs of the detection step, but also suppresses spontaneous detection hypotheses, which are not stable over several frames, such as certain constellations in inhomogeneous regions (sun shining through trees might look like a traffic light, but is usually not stable over several frames).
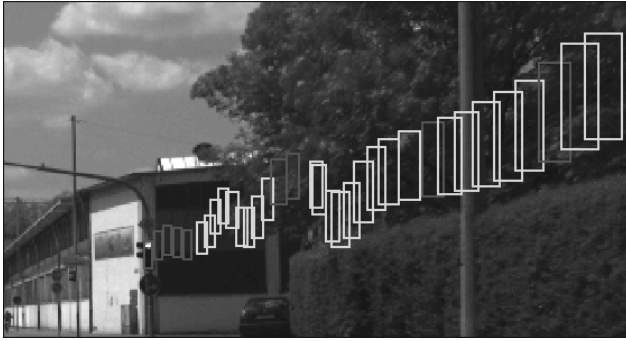

Fig. 2: Detection and tracking of a traffic signal. Bright boxes are detected and dark boxes are not available to the tracker.

Tracking may also be used to support the detection process at certain regions which had a positive detection in the last frame, or to track previously recognized traffic signals by the use of correlation tracking [4]. In the urban scenario we must always expect (unless using differential GPS and enhanced digital maps), that traffic signals may appear almost anywhere in the image, not only at the focus of expansion or at the image borders, since any object might temporarily be hidden by traffic participants or other obstacles.

C. Classifier

The final stage is a classifier, which consecutively evaluates each element of a track. The overall decision is made by soft majority voting, taking into account the discrimination values of each frame. A preliminary decision can be made as soon as the collected probability mass is large enough. The challenging task of the classifier module is to reject false candidates as robustly as possible. While for a color camera the status of the traffic signal is known implicitly, for gray value cameras the status has to be classified as well.

Fig. 3: Samples of red, yellow and green traffic signals as well as false hypotheses, all used for classifier adaptation.

For classification we use feed-forward neural networks with receptive fields [5]. For adaptation of the classifier bootstrapping techniques are applied [3], which automatically enlarge the training set with false alarms, thus improving the reject capabilities of the single frame classifier. The soft majority voting further improves the performance of the single frame classifier significantly.

D. Optional Sensors

In modern cars vehicle data (velocity, yaw rate) are available on the CAN-bus, which can be used to enhance the tracking considerably. Using a 3D-from-motion algorithm the real world coordinates of the traffic signal relative to the camera position can be computed after a few frames. The distance to the traffic signal can be estimated rather accurately, as long as the traffic signal is not lined up directly with the car motion vector – see figure 4. The distance accuracy is about 1 m depending on the setting and the used yaw sensor. The estimation of the height is usually unreliable (and therefore useless), since cars pitch during braking, acceleration, and when driving through ditches, and pitch sensors are not standard equipment.

A major impact on the improvement of the traffic signal recognition system comes from enhanced map information in combination with GPS or differential GPS – see also section III. Rough position information (10 m accuracy) of the car and the traffic signals allows to start the system only

in front of equipped intersections and no false alarms will occur in between (reducing the number of false alarms by a factor of 20 in our experiments). With high accuracy positioning systems and maps the detector can even be narrowed to image regions where to look for traffic signals. Using 1 m GPS accuracy and 1° heading uncertainty another factor of about 5 is gained on the rejection rate. Furthermore, if information about the lane, the vehicle is currently driving in, is available, the assignment of the different traffic signals to the according lane is done easily using the enhanced map information.
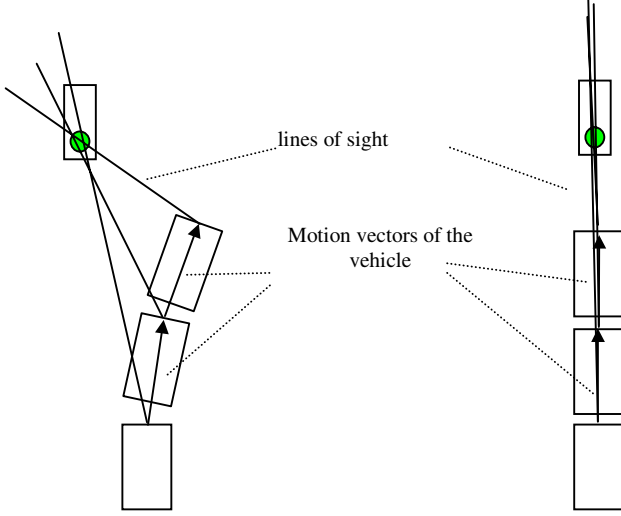

Fig. 4:   3D-from-motion: standard situation is depicted on the left side, no reliable distance estimation is possible for the situation on the right side.

Using a second camera allows the use of stereo algorithms to measure the real world distance and size of the detected objects. The number of candidates for traffic signals is reduced by a factor of about 10 in our experiments by limiting the possible real size of traffic lights. Additionally the tracker is improved by including distance and size information. The cameras need to be calibrated and to have synchronized triggers. The distance can be measured with an accuracy of approximately 50 cm. In combination with GPS and digital map information stereo systems have an even higher impact, since rejection can be done more strictly.

III. COLOR DETECTOR

In [1] we proposed a traffic signal recognition system based on a color camera. The major challenge at that time was the low dynamic range of color sensors. In order to verify the active traffic light by classification, two inactive traffic lights have to be visible above/below it. To discern this region from the background, the exposure time had to be long enough, which forced the active traffic lights into saturation hence producing white instead of the signal colors. Using high dynamic range CMOS sensors the whole scene is discernable and we still have detectable signal colors.

The color detector uses a modified Gaussian-distribution classifier [6, 7] to determine the class of each pixel (red, yellow, green or background). The classifier is trained with several thousand example pixels from labeled images. The image resulting from the classification process is blurred and then segmented by the use of a connected component analysis [8]. Scattered noise pixels are filtered out and regions containing the lights of traffic signals are detected.
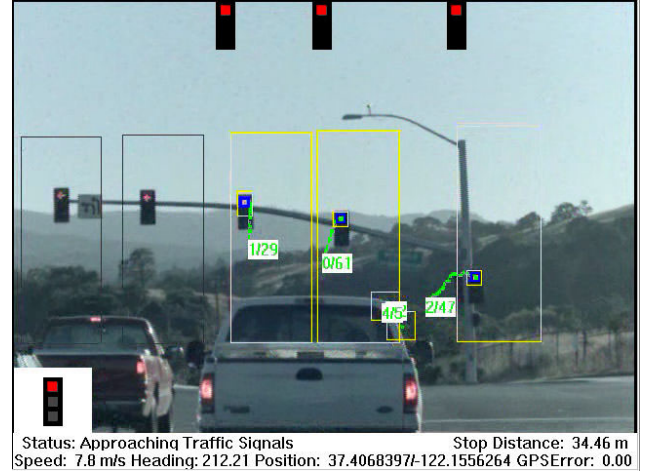

Fig. 5:   Output of the IVI-Edmap system, detector is running in the three boxes on the right side only.

A system using this detector, differential GPS, enhanced digital maps and 3D-from-motion is in use in the IVI-Edmap research project at DaimlerChrysler USA, running at 25 Hz on a 3 GHz Pentium IV. It recognizes over 95% of all traffic signals from up to 100 m distance, generating less than one false positive per hour for urban driving. The sensor is a one chip 10 bit VGA-resolution High Dynamic Range Camera (HDRC) with a 16 mm lens.

IV. GRAY VALUE DETECTORS

The sensor for the gray value system is a 12 bit VGA resolution camera with a 12 mm lens. The system is running on a 1.2 GHz Pentium III computer.

For the systems described in sections IV-A and IV-B, the classifier has to verify the complete traffic signal, without knowing which light is active and detected. Therefore three different cut-outs are generated for each detected light – the cut-out is centered around the light for assuming a yellow light, and shifted downwards/upwards for assuming red/green lights. The hypothesis is rejected if all three cut-outs yield discriminate values below a threshold. Otherwise the winner is taken for soft majority voting over the complete track. The adaptation of the classifier is otherwise identically to the color camera case.

A. Shape-Based Detector

The lights in traffic signals are round, except directional signals for left/right turning lanes (in Germany even the stop-state is depicted by a black arrow on a red light, while the go-state is just a green arrow; in the USA colored arrows are used for all signal states). Since we already implemented a highly efficient circle finder for traffic signs [3], we used this system here as well. The circle finder is based on a generalized Hough transform using Sobel based gradient directions. Traffic lights are detected if the diameter is larger than 6 pixels. In order to detect even slightly distorted lights, the quality threshold for the circle is reduced, generating however more false hypotheses as known from the traffic sign recognition [3]. Nevertheless, we reached recognition rates for the complete system of over 80 % with less than one false alarm per 400 frames. Due to the highly optimized version of the Hough transform the whole system runs at over 15 frames per second.
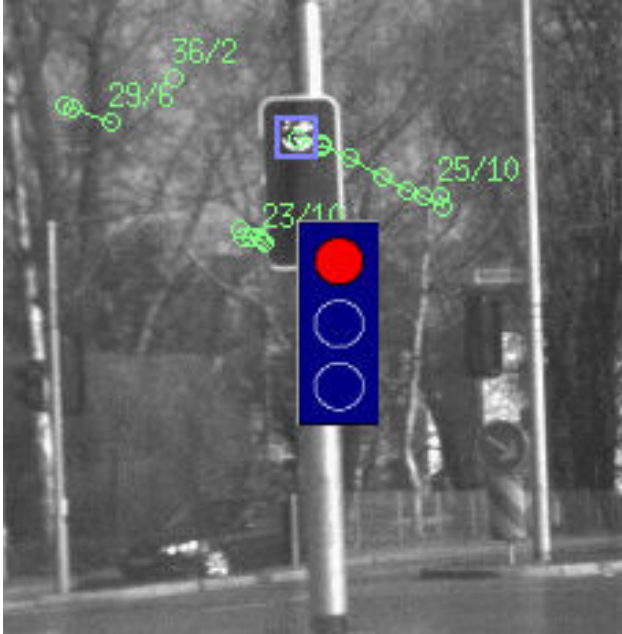

Fig. 6:   Results of Hough transform circle detector.

The major challenge for a shape-based detector are the turning signals (direction arrows), which will require different shape finders and thus slow down the system significantly. Instead of implementing new shape detectors based on edge features, we decided in favor of template based matched filters.

B. Matched Filter Algorithm

Typical for any traffic light is a brighter spot surrounded by a darker box (this holds true even for arrows), which leads to a template as seen in figure 7 on the left side. In order to match the template in different sizes (from 4 to 20 pixels in 8 steps) at all possible positions a highly computational efficient matching approach is necessary.

Fig. 7:   Coarse box and directional matching template.

If an integral image [9] is used, which is computed only once for each frame, the matched filter needs only 8 add/sub operations for each match. Of the about 2.5 million regions tested per image up to 400 candidates with high matching scores are further processed by the directional template in figure 7 on the right side, which tests whether the mean values in each of the 4 surrounding quadrants are lower than in the center. This more elaborate second template suppresses 80% of the given candidates, leading to 20 - 80 regions for the tracker.

The higher false positive rate of the matched filter compared to the shape-based detector is compensated by the tracker and the neural net. The complete recognition system runs at approximately 10 Hz with a detection rate larger than 85 % and less than one false positive per 300 frames.
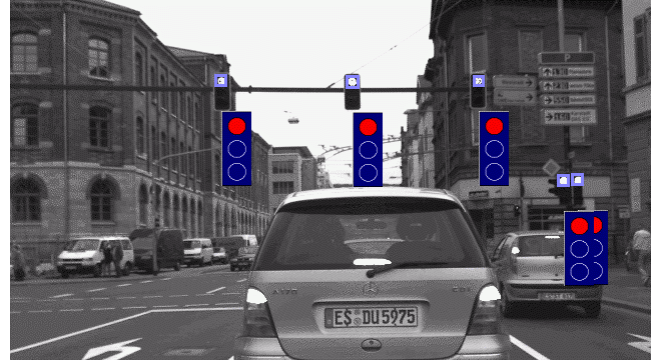

Fig. 8:   System results using the matched filter detector.

C. Detection using Cascade Classifiers

Since we succeeded to match special templates based on the integral image over the complete image in real time, the question arises whether complete-object classifiers can be applied at all positions with different sizes over the complete image and still fulfill the real time requirements. In [9] a method is proposed, which is based on the integral image (as the matched filter) and gains its speed by a pipeline of classifiers, starting with a very fast one with only very few features and increasing the number of features successively through the pipeline. The convincing idea is, that each classifier in the pipeline rejects as many candidates as possible without loosing any object searched for. Only candidates not rejected, consisting of objects and false alarms, are

passed to the next classifier in the pipeline. The important task, to find features that are as discriminative as possible, is solved automatically with the AdaBoost algorithm combining weak learners (= features) to a strong classifier. The method was applied successfully in [9] to face detection.
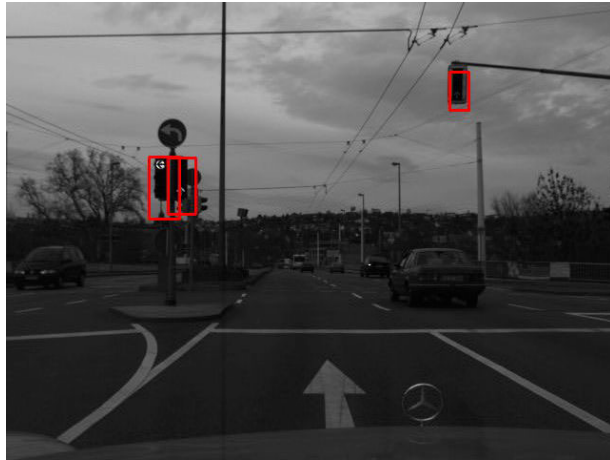

Fig. 9:   Detection results of a cascade classifier.

We apply this method using our data bases of traffic signals in a rather standard fashion. We used about 10 different scalings between 20 and 120 pixels and normalized the brightness of the regions. The resulting rectangles are clustered to represent one box per object only (see figure 9). Since the verification step is implicitly done in the last steps of the pipeline, the method yields a lower number of false positives. Due to the large depth of the pipeline (10 or more boosting classifiers) the system is still too slow for the real time requirement (cycle time about 500 msec per frame). An important advantage is, however, that different traffic lights (round, arrow, ...) can be integrated automatically without changing the system.

V. SUMMARY AND CONCLUSIONS

In the following table our experiences with different settings are summarized:

Detector	System Performance	Challenges
color + diff. GPS/map	25 Hz >95% recognized 1 fpos/hour	costly color sensor
shape based matched filter + GPS + GPS + stereo	80% recognized / 15 Hz 85% recognized / 10 Hz 1-4 fpos/minute 1 fpos/5 minutes 1 fpos/30 minutes	few directional lights not all directional lights
cascade classifiers	90% recognized/2 Hz 1-2 fpos/minute	not real time, recognition rates taken offline at simulated 10 Hz

Tab. 1:   Performance summary for urban environment: all systems with vehicle data; fpos = false positives/alarms.

The color camera system is capable of becoming a valuable help in urban traffic by directing the driver's attention to traffic signals, as soon as color cameras are affordable and the traffic signals are registered in the enhanced digital maps. Precise location information and/or a lane detection algorithm will perfect the approach to a robust and reliable system (comparable to transponder systems, which have significant additional infrastructure costs).

Without color information the detection step is still very critical, but might be solved by a smart combination of the cascade classifiers and the matched filter algorithm – the recognition performance of a complete-object detector/classifier and the real time capability offered by designed matched filters.

REFERENCES

[1] U. Franke, D. Gavrila, S. Goerzig, F. Lindner, F. Paetzold, C. Woehler:
"Autonomous Driving Goes Downtown"
IEEE Intelligent Systems, vol. 13, no. 6, pp. 40-48, 1998.
[2] http://www.invent-online.de; project VAS.
[3] F. Lindner, U. Kressel, C. Woehler, A. Linz:
"Hypothesis verification based on classification at unequal error rates"
International Conference on Artificial Neural Networks, pp. 874-879, Edinburgh, 1999.
[4] S. Gehrig, S. Wagner, U. Franke:
"System Architecture for an Intersection Assistant Fusing Image Map and GPS Information"
Intelligent Vehicles Conference, pp. 144-148, 2003.
[5] C. Woehler, J. Anlauf:
"Real-time object recognition on image sequences with the adaptable time delay neural network algorithm - applications for autonomous vehicles"
Image and Vision Computing, vol. 19, no. 9-10, pp. 593-618, 2001.
[6] S.Kaelberer:
"Detektion von Ampeln fuer ein Verkehrszeichenerkennungssystem"
Diploma thesis, Universitaet Ulm, 2003.
[7] R. Duda, P.Hart, D. Stork:
"Pattern Classification"
John Wiley & Sons, 2001.
[8] E.Mandler, M. Oberlaender:
"One Pass Encoding Of Connected Components in Multi-Valued Images"
IEEE Int. Conf. on Pattern Recognition, pp. 64-69, 1990.
[9] P. Viola, M. Jones:
"Robust Real-Time Object Detection"
Technical Report CRL 2001/01
Compaq Cambridge Research Laboratory, 2001.
[10] S. Wender, O. Loehlein:
"A Cascade Detector Approach Applied to Vehicle Occupant Monitoring with an Omni-directional Camera"
IEEE Intelligent Vehicles Symposium, Parma, 2004.