Real-time Vehicle Signal Lights Recognition with HDR Camera

Jian-Gang Wang, Lubing Zhou, Zhiwei Song and Miaolong Yuan

Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR)
1 Fusionopolis Way, #21-01 Connexis (South Tower), Singapore 138632

jgwang@i2r.a-star.edu.sg

***Abstract*—Light signal recognition is one of the key issues in autonomous vehicle. Unlike normal object detection and recognition, which can be done by using different sensors, light signal recognition is naturally a computer vision problem. Although commercialized ADAS (Advanced Driving Assistance System) products, such as mobileye, could be used for rear-end collision warning, a cost-effective approach is still needed. In this paper, we have developed a novel two-stage approach to detect vehicles and recognize signal lights from a single image in real-time. Distinct from current state-of-the-art light recognition algorithms, we adopted a HDR (High Dynamic Range) camera instead of a color camera. Taking advantage of the different dynamic range of a HDR camera, the detection of a vehicle and the recognition of the signal light have been done in bright and dark channels, respectively. Furthermore, unlike previous approaches where pair taillight has to be extracted explicitly, we use detected vehicle region instead. On a large database, "Lights Patterns" (LPs) are learned by a multi-layer perception neural network. The robustness and efficiency of the proposed approach has been verified by the experimental results conducted on some real on-road videos.**

Keywords— Signal lights recognition; Autonomous vehicle; Deep learning; Real-time; Computer vision.

I. Introduction

One of the capabilities of an autonomous vehicle is that it should be able to understand the intention and behavior of the vehicle ahead and take property navigation to prevent rear-end collisions and accidents. The intention information of a driver comes mainly from the vehicle's light signals. To automatically recognize the light signals, some technologies such as color image processing and blob algorithm have been explored and it is naturally a computer vision problem. The light signals include brake, reverse, left turn, right turn, parking, headlights etc. As the recognition of different light signals could be done in a similar way, we will discuss only the brake-lights recognition problem in this paper.

There are a few approaches in the past decade to detect and recognize signal lights. The existing approaches can be classified into two categories based on the information being used: (1) temporal information [1-3]; (2) single image [4-7]. Most of them detect brake lights using red color features and pair taillights according to the symmetry of the vehicle rear lights. For instance, a hierarchical algorithm to detect vehicle and rear-lights in the daytime is developed by Cui et al [7]. In their approach, vehicle is detected by the DPM (Deformable Part Model) [8] and then red light candidates can be found by clustering pixels in the HSV color space. They adopted a sparse dictionary learning to recognize brake-lights after pairing taillights. The DPM is too slow for real-time application and the pairing taillights could result in failure because of the possible rear-lights occlusion. Furthermore, the detection of the rear-lights could be affected by the noise from the urban road environment, e.g. traffic lights, streetlight.

In this paper, we propose a fast approach to robustly detect rear-lights from a single image. It is a 'two-stage" approach: vehicle detection and light signals recognition. In the first stage, one prefers to detect vehicles from a bright image because of the vehicles' large color and shape variation. In the second stage, considering the signals to be recognized come from lights, a clear and dark background should be much more useful to achieve a better recognition rate. This motivates us to adopt HDR camera.

HDR imaging is a method that aims to add more "dynamic range" to photographs, where dynamic range is the ratio of light to dark in a photograph. Instead of just taking one photo, HDR camera provides two synchronized channels taken at different exposure. The corresponding pixels between the two channels can be found easily. We have published a brake-light recognition approach [9] in which the brake-lights are recognized based on a color camera. In this paper, we extend our previous approach in two aspects. Furthermore, we also adopt deep learning to detect the vehicle from an image, allowing us to take advantage of the latest deep learning object method. Compared to the previous algorithm in [9], the method presented in this paper can achieve faster speed and higher accuracy.

Different from the existing brake-lights recognition approaches in which the left and right rear-lights have to be extracted explicitly, an appearance-based deep learning algorithm is proposed in this paper. In order to recognize light signals of a vehicle, the regions in the dark channel, corresponding to the detected vehicle regions in the bright channel, are passed to the light signal recognition module. The advantages of our approach are three-fold: (1) The recognition of the light signal is much more robust than the state-of-the-art because a HDR camera is used; (2) Besides the left and right rear lights, the middle top light is taken into consideration. (3) The occlusion problem could be solved in somewhat extent because our approach does not need to extract left and right signal lights explicitly.

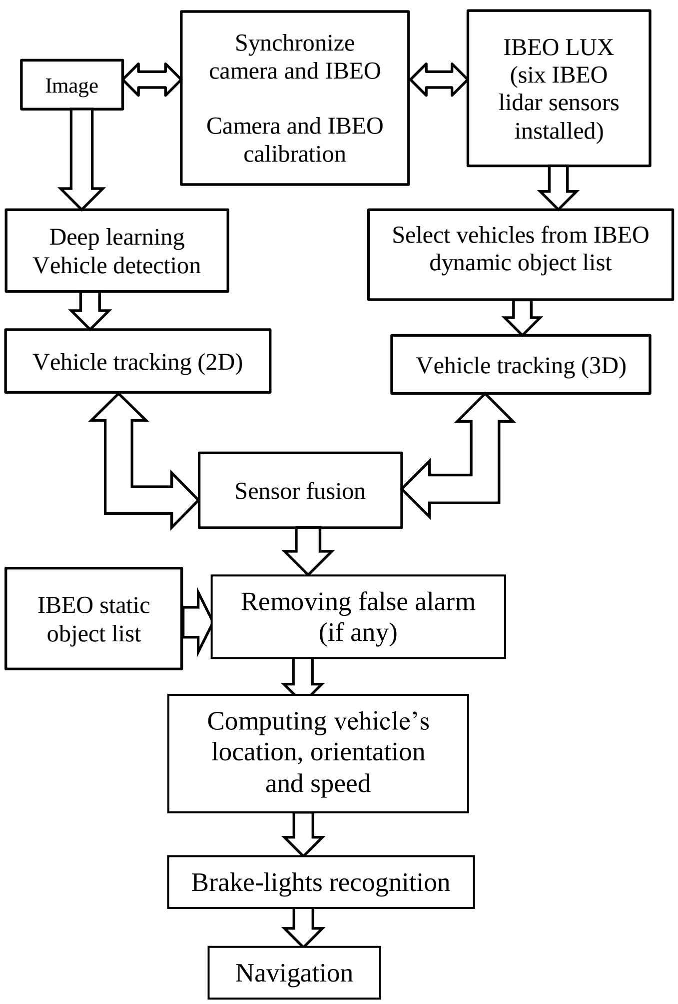

Figure 1. Proposed two-stage brake-lights recognition system

The flowchart of the proposed approach is shown in Figure 1. The algorithms have been implemented in C++ and integrated into our autonomous vehicle (IIR AV) [10] via chromosome (XME) operating system. IBEO LUX [11] is installed in the AV and sensor fusion has been developed to improve the vehicle detection performance.

II. MULTIMODAL VEHICLE DETECTION

Vehicle detection or general object detection has been extensively studied in the computer vision and intelligent transportation system community. A good survey about this topic can be found in [12]. It is challenging to detect vehicles on real road scenery because of the complexity of the background. The state-of-the-art vehicle detection is the fusion of 2D and 3D sensors. In this paper, vehicles are detected by combining a commercialized multi-layer lidar sensor (IBEO LUX [11]) and a HDR camera.

A. *Vehicle detection*

The state-of-the-art object detection from an image is based on deep learning technology. Similar idea is proposed for this purpose: hypothesize bounding boxes, resample pixels or features for each box, and apply a high-quality classifier. This means that some candidate bounding boxes have to be found before recognition. The computation time for looking for the candidate is light, however, the number of the candidates is high, for instance 2000. This makes the detection too slow to meet the real-time applications. By reducing the number of the candidates, from 2000 to 300, and sharing features within the network, Faster RCNN [13], one of the examples, can operate at only 7 frames per second (FPS).

There are few deep learning object detection approaches which can achieve high frame rates while providing a comparable accuracy with the state-of-the-art algorithms. Examples are the darknet [14] and SSD (single-shot detector) [15]. Darknet formulates the detection as a regression problem and achieve a fast frame rate by accessing the image only once. The algorithms do not require bounding boxes. In this paper, we adopt SSD to detect vehicle from a signal image.

The SSD discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default box and produces adjustments to the box to better match the object shape. Additionally, the network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes. For 300×300 input, SSD achieves 72.1% mAP on VOC2007 test at 58 FPS on a Nvidia Titan X and for 500×500 input, SSD achieves 75.1% mAP, outperforming a comparable state of the art Faster R-CNN model.

B. *Sensor fusion*

Lidar and camera can be fused to make a final decision as this tow kinds of sensors compensate for each other in terms of feature. Lidar provides depth but lack texture. In contrast, camera is rich in texture information. We combine Lidar and camera in this paper to improve the detection accuracy.

We will focus on vehicle detection and signal lights recognition from image. The interest reader can refer [9] for the details of the sensor fusion.

In Figure 3, we show some examples of vehicle detection. The results are obtained by fusing IBEO (both dynamic and static object lists) and vision.

Figure 3. Vehicle detection by fusing IBEO LUX and vision

III. Brake-lights recognition

Deep learning has recently achieved state-of-the-art performance on a number of image recognition benchmarks, including ILSVRC-2012 [16]. Deep learning technology has been adopted in this paper to recognize brake-lights from image because light-patterns (LPs) of the brake-lights can be learned well from the HDR dark channel. The clear and dark background of the dark channel image makes the lights recognition robustly. Another advantage of proposed approach is that the occlusion problem can be overcome to some extent because appearance image rather than a pair of rear-lights is used. The training samples for occlusion cases are included. Furthermore, the middle brake-lights (located at the top of the rear window) belong to the LP besides the left and right rear-lights. To the best of our knowledge, there is little literature in which this middle rear brake-lights is taken into consideration. The reason it is not being used in previous approaches could be that the middle light is relatively darker than left and right rear-lights and extracting it from the image is not easy.

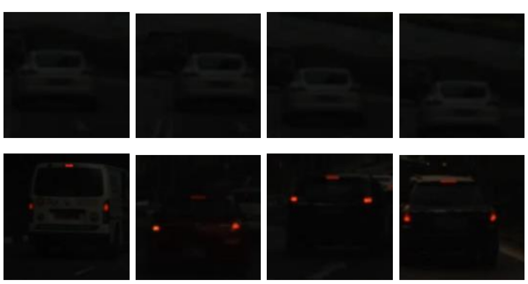

Figure 4. Some training samples for “normal” (first row) and “brake” (second row)

Feature extraction is an important issue in appearance-based recognition. The objective of the feature extraction is to find a representation for machine learning algorithms. Deep learning, state-of-the-art machine learning, has gained impressive performance in the domain of object recognition [17]. Different from conventional hand-crafted features, deep learning has the ability to extract self-learning features from a huge size data. For instance, ImageNet dataset [16] contains 1000-class samples to be used for a classification task; the deep models pre-trained on ImageNet have shown excellent accuracy for some challenging problems, e.g. object detection [18] or semantic segmentation [19].

Open sources, e.g. Caffe [20], provide publicly available deep convolutional neural networks, e.g. VGG-16 [21] or AlexNet model [22], for different deep learning applications. In order to achieve a fast algorithm, the AlexNet model, a relatively simple, eight-layered network, is adopted in this paper.

IV. Experimental Results

A. Data collection

The experimental data in this paper was collected on the real-road scenery for two months. The variances of the weather condition, time of the day, have been considered for this data collection.

The detected vehicles (represented as rectangles on images) are manually labeled as “brake” or “normal” categories. About 3700 “brake” samples (positive) and 1900 “normal” samples (negative) are annotated (called seed samples) from about 300 videos. The image has a resolution of 1600 × 1200. A total of one million training samples are generated from the seed samples. The ratio between the positive and negative training samples has been found to be important in achieving a high accuracy classifier. In our experiments, a total of half million positive samples as well as half million negative samples are generated, respectively.

The method to generate new samples is described as follows. Shifting the center of the vehicle rectangle with a uniform random distribution from -0.2 to 0.2 times of the rectangle’s width or height, and then resizing the rectangle with a uniform random distribution from 1 to 1.2 times. The new samples are resized to a standard size required by the Caffe. For AlexNet model used in this paper, the sample size is 224 × 224. In figure 4, some new samples obtained using the above method are shown.

B. Classifier training

An AlexNet model is finetuned using the dataset described in last section. The weights provided by BVLC AlexNet model [22] is used as the initial weights for the fine-tuning.

A number of brake-lights recognition results are given in Figure 5 to 7. The cars recognized as “brake lights” are shown in red, otherwise shown in green.

TABLE I. Time costs for processing one frame

Function	Vehicle detection	Brake-lights recognition	Total
Time (ms)	2	12	14

The algorithms proposed in this paper are implemented in C++. It runs on a desktop computer, with a Titan Black graphics card, 64-bit, 16GB memory. The vehicle detection and brake-lights recognition can run in about 65-70 fps depends on the number of the vehicles in an image. The average time costs for one frame are listed in Table 1.

Figure 5. Left: “normal” (green); right: “brake” (red)

Figure 6. Left: A lorry vehicle without middle brake-lights; Right: Vehicles with/ without middle brake-lights.

Figure 7. The right vehicle on the right can be recognized even its right rear light is fully occluded.

It is hard for us to quantitatively compare the proposed approach with the state-of-the-art approaches because no public benchmark database is available for brake-lights recognition. Nevertheless, the proposed approach has been tested (ten-fold cross-validation) on our own database. The average accuracy is found to be 97.5%, much better than the one, 89%, obtained by using normal color image [9]. A similar evaluation has been conducted for vehicle detection. When only rear views are taken into consideration, the average accuracy is found to be 99.5%, better than the previous one, 99%, obtained by darknet [14].

The algorithms developed in this paper have been integrated into an autonomous vehicle, iiRAV [10]. The demonstrations on real road, including vehicle following, obstacle avoidance, etc., have shown that both the accuracy and the speed given in this paper are satisfied with the navigation requirement.

V. Conclusion and Future Work

In this paper, we have proposed a real-time two-stage appearance-based deep learning method to recognition brake-lights from a single image. Different from the state-of-the-art approach, HDR camera is used instead of normal color camera. The clear background of HDR dark channel makes the light signal recognition much more reliable than the one based on color image. In addition, the state-of-the-art deep learning technique has been adopted to detect vehicle and recognize signal lights. The experimental results on a large database have shown that the brake-lights recognition accuracy is much higher than the one obtained by using a normal camera.

The experiments on other kind of light signals including left turn, right turn will be evaluated in the near future. The recognition of the brake-lights in the night time could be done by developing video analysis.

References

[1] 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, no. 2, pp. 453– 462, 2010.

[2] M. Casares, A. Almagambetov, and S. Velipasalar, “A robust algorithm for the detection of vehicle turn signals and brake lights,” in IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance, 2012, pp. 386–391.

[3] A. Almagambetov, M. Casares, and S. Velipasalar, “Autonomous tracking of vehicle rear lights and detection of brakes and turn signals,” in IEEE Symposium on Computational Intelligence for Security and Defence Applications (CISDA), 2012, pp. 1–7.

[4] Z.-Y. Cui , S.-W. Yang, and H.-M. Tsai, “A Vision-Based Hierarchical Framework for Autonomous Front-Vehicle Taillights Detection and Signal Recognition,” in 2015 IEEE 18th International Conference on Intelligent Transportation Systems, pp. 931-937.

[5] P. Thammakaroon and P. Tangamchit, “Predictive brake warning at night using taillight characteristic,” in IEEE International Symposium on Industrial Electronics, 2009, pp. 217–221.

[6] Q. Ming and K.-H. Jo, “Vehicle detection using tail light segmentation,” in 6th IEEE International Forum on Strategic Technology (IFOST), 2011, vol. 2, pp. 729–732.

[7] S. Nagumo, H. Hasegawa, and N. Okamoto, “Extraction of forward vehicles by front-mounted camera using brightness information,” in IEEE Canadian Conference on Electrical and Computer Engineering, 2003, vol. 2., 2003, pp. 1243–1246.

[8] P. F. Felzenszwalb, R. B. Girshick, and D. McAllester, “Cascade object detection with deformable part models,” in IEEE conference on Computer vision and pattern recognition, 2010, pp. 2241–2248.

[9] J.-G. Wang, L.-B. Zhou, Y. Pan, S. Lee, Z.-W. Song, B. S. Han, V. Billy, Appearance-based Brake-Lights recognition using deep learning and vehicle detection, Intelligent Vehicles Symposium 2016: 815-820.

[10] IIRAV, http://www.i2r.a-star.edu.sg/autonomousvehicle/

[11] IBEO: http://www.ibeo-as.com/

[12] S. Sivaraman, M. M. Trivedi, “Looking at Vehicles on the Road: A Survey of Vision-Based Vehicle Detection, Tracking, and Behavior Analysis,” IEEE Transactions on Intelligent Transportation Systems, 14(4):1-23, 2013.

[13]S.-Q. Ren, K.-M. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” arXiv:1506.01497 [cs.CV].

[14] arXiv: 1506.02640 [cs.CV], http://pjreddie.com/darknet/yolo/

[15] SSD: Single Shot MultiBox Detector, https://arxiv.org/abs/1512.02325

[16] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al., “Imagenet large scale visual recognition challenge,” arXiv:1409.0575, 2014.

[17] A. Krizhevsky, I. Sutskever, G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” NIPS, 2012.

[18] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, “Scalable Object Detection using Deep Neural Networks,” arXiv: 1312.2249[cs.CV].

[19] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” arXiv.1411.4038 [cs.CV].

[20] http://caffe.berkeleyvision.org/

[21] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv:1409.1556 [cs.CV].

[22] https://github.com/BVLC/caffe/tree/master/models/bvlc_alexnet