Expert Systems With Applications 88 (2017) 290–304

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Traffic light recognition exploiting map and localization at every stage

Chulhoon Jang[a], Sungjin Cho[a], Sangwoo Jeong[a], Jae Kyu Suhr[b], Ho Gi Jung[b], Myoungho Sunwoo[a,*]

[a] Department of Automotive Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 04763, Republic of Korea
[b] Major of Information & Communications Engineering, Korea National University of Transportation, 50 Daehak-ro, Chungju-si, Chungbuk 27469, Republic of Korea

ARTICLE INFO

Article history:
Received 5 February 2017
Revised 30 June 2017
Accepted 5 July 2017
Available online 12 July 2017

Keywords:
Traffic light recognition
Localization and digital map
Intelligent vehicles
Intelligent transportation systems
Computer vision
Object detection

ABSTRACT

Traffic light recognition is being intensively researched for the purpose of reducing traffic accidents at intersections and realizing autonomous driving. However, conventional vision-based approaches have several limitations due to full image scanning, always-on operation, various different types of traffic lights, and complex driving environments. In particular, it might be impossible to recognize a relevant traffic light among multiple traffic lights at multiple intersections. To overcome such limitations, we propose an effective architecture that integrates a vision system with an accurate positioning system and an extended digital map. The recognition process is divided into four stages and we suggest an extended methodology for each stage. These stages are: ROI generation, detection, classification, and tracking. The 3D positions of traffic lights and slope information obtained from an extended digital map enable ROIs to be generated accurately, even on slanted roads, while independent design and implementation of individual recognition modules for detection and classification allow for selection according to the type of traffic light face. Such a modular architecture gives the system simplicity, flexibility, and maintainable algorithms. In addition, adaptive tracking that exploits the distance to traffic lights allows for seamless state estimation through smooth data association when measurements change from long to short ranges. Evaluation of the proposed system occurred at six test sites and utilized two different types of traffic lights, seven states, sloped roads, and various environmental complexities. The experimental results show that the proposed system can recognize traffic lights with 98.68% precision, 92.73% recall, and 95.52% accuracy in the 10.02–81.21 m range.

1. Introduction

Recently, the traffic light recognition (TLR) problem has become an important research focus for driver assistant systems (DAS) and autonomous driving. Perception of traffic lights (TL) at intersections and crosswalks is a necessary function for compliance with traffic regulations and to prevent fatal road accidents (De Charette & Nashashibi, 2009a; Fairfield & Urmson, 2011; Gomez, Alencar, Prado, Osorio, & Wolf, 2014; Jensen, Philipsen, Møgelmose, Moeslund, & Trivedi, 2016; Kim, Shin, Kuk, Park, & Jung, 2013; Levinson, Askeland, Dolson, & Thrun, 2011; Yu, Huang, & Lang, 2010). There are two approaches to TLR: a communication-based method and a vision-based method. The communication-based method utilizes wireless communications between the vehicle and surrounding infrastructure, in addition to providing TL states to surrounding vehicles (Bazzi, Zanella, & Masini, 2016; Ferreira & d'Orey, 2012; Florin & Olariu, 2015; Younes & Boukerche, 2016). This method may be the most faithful way to deliver traffic information, but it requires substantial investment for the installation of wireless devices in all traffic lights and vehicles (Fairfield & Urmson, 2011; Jensen et al., 2016; Levinson et al., 2011; Yu et al., 2010). On the other hand, the vision-based method utilizes a camera installed on a windshield for TLR and has considerable potential to increase functionality for customers. However, this method can also be easily affected by disturbances in the surrounding area such as objects with similar shapes and colors.

Conventional vision-based approaches have been researched in various ways (Diaz, Cerri, Pirlo, Ferrer, & Impedovo, 2015; Jensen et al., 2016). These approaches have several limitations in DAS or autonomous driving as they attempt to meet stringent requirements. First, the camera requires a wide field of view (FOV) and high-resolution to cover short and long ranges for TLR. The high-resolution image increases the amount of information and

Abbreviations: TLR, traffic light recognition; TL, traffic light; FOV, field of view; DAS, driver assistant system; ROI, region of interest; LIDAR, light detection and ranging; HOG, histogram of oriented gradients; SVM, support vector machine.

* Corresponding author.

E-mail addresses: chulhoonjang@gmail.com (C. Jang), sungjincho215@gmail.com (S. Cho), sangwoojeong0318@gmail.com (S. Jeong), jksuhr@gmail.com (J.K. Suhr), hogijung@ut.ac.kr (H.G. Jung), msunwoo@hanyang.ac.kr (M. Sunwoo).

http://dx.doi.org/10.1016/j.eswa.2017.07.003
0957-4174/

requires large amounts of computing power. Second, many false positives may be generated since a TL is a very tiny object with few distinctive features; in particular, it is hard to distinguish between TLs and the brake lights of preceding vehicles on slanted roads, or to identify TLs against complex backgrounds with lights and colored signs (Gomez et al., 2014; Jensen et al., 2016; Omachi & Omachi, 2010). Third, the various compositions of TLs lead to sophisticated problems of classification and decision making: red, green, yellow, and arrows can be combined in many different ways (Jensen et al., 2016). And last, recognition of TLs in multiple lanes and multiple TL conditions is a much more challenging problem than simple recognition of TL states (Jensen et al., 2016).

Various methods to utilize localization and map information for TLR have been proposed as this approach uses prior knowledge contained within a map to make online perception simpler and more efficient (Barnes, Maddern, & Posner, 2015; Fairfield & Urmson, 2011; John, Yoneda, Qi, Liu, & Mita, 2014; Levinson et al., 2011; Lindner, Kressel, & Kaelberer, 2004; Tae-Hyun, In-Hak, & Seong-Ik, 2006; Ziegler et al., 2014). In other words, prior knowledge leads to improvements in the accuracy of recognition and reduces algorithm complexity. For instance, the recognition algorithm does not need to operate continuously as perception begins only when the distance to the facing TL is within a certain threshold. As a result, resources can be saved and efficiency enhanced.

This paper introduces a method to extend localization and map information to all stages of TLR and improve recognition performance. The TLR system is divided into four stages: ROI generation, detection, classification, and tracking. ROI generation based on the 3D positions of TLs is a useful way to reduce scan areas within images, but conventional approaches do not take into account slope conditions and so may fail to recognize TLs located on slanted roads. The proposed ROI generation method not only exploits the 3D position of TLs, but also exploits slope information to create precise candidate regions even in uphill or downhill conditions. Furthermore, it is difficult to design a generalized recognition module due to different conventions between countries and non-standardized TLs; thus, this paper introduces a modular architecture to selectively apply specific detectors and classifiers according to the TL type of facing TLs. Lastly, nearest neighbor filters are commonly used for tracking, but track loss or unnecessary track generation might occur with fixed association parameters. As such, this paper proposes an adaptive approach using distance to TL to enhance seamless tracking from long to short range.

This paper provides the following contributions

(1) It presents a method that reliably compensates for road slopes to precisely generate ROIs by applying linear interpolation between constant slope regions and flat regions.
(2) It suggests a modular architecture for TLR. Individual recognition modules for all TL types are trained separately offline and selected according to the online retrieved TL type.
(3) It introduces adaptive data association to compensate for perspective deformation by using distance to TL.

This paper evaluates the proposed methods using on-road experimental data, and the results of a comparison with a standalone vision method show that extending localization and map information to an advanced perception system can secure computational efficiency, reduce false positives, and improve the recognition rate.

2. Related works

Much research has been conducted on TLR, along with the recent introduction of survey papers (Diaz et al., 2015; Jensen et al., 2016). The conventional approach is a pure vision based approach (Almagambetov, Velipasalar, & Baitassova, 2015; Angin, Bhargava, & Helal, 2010; Anh, Ramanandan, Anning, Farrell, & Barth, 2012; Borrmann et al., 2014; Cai, Gu, & Li, 2012; Chen, Shi, & Huang, 2015; Chin-Lun, Shu-Wen, & Jyh, 2009; Chiu, Chen, & Hsieh, 2014; Chung, Wang, & Chen, 2002; De Charette & Nashashibi, 2009a, 2009b; Diaz-Cabrera & Cerri, 2013; Diaz-Cabrera, Cerri, & Medici, 2015; Diaz-Cabrera, Cerri, & Sanchez-Medina, 2012; Fan, Lin, & Yang, 2012; Gomez et al., 2014; Jang, Kim, Kim, Lee, & Sunwoo, 2014; Jensen et al., 2015; Jianhua, 2015; Jianwei et al., 2010; Jongwon, Byung Tae, & In So, 2013; Kim et al., 2013; Kim, Kim, & Yang, 2007; Li, Cai, Gu, & Yan, 2011; Michael & Schlipsing, 2015; Nienhuser, Drescher, & Zollner, 2010; Omachi & Omachi, 2009, 2010; Philipsen, Jensen, Mogelmose, Moeslund, & Trivedi, 2015; Philipsen, Jensen, Trivedi, Mogelmose, & Moeslund, 2015; Shi, Zou, & Zhang, 2015; Sooksatra & Kondo, 2014; Trehard, Pollard, Bradai, & Nashashibi, 2014; Xuemei, Yanmin, Minglu, & Qian, 2012; Yehu, Ozguner, Redmill, & Jilin, 2009; Ying, Chen, Gao, & Xiong, 2013; Yu et al., 2010; Zhang, Fu, Yang, & Wang, 2014; Zixing, Yi, & Mingqin, 2012; Zong & Chen, 2014), but there has been less work on utilizing localization and map information for TLR (Barnes et al., 2015; Fairfield & Urmson, 2011; John et al., 2014; Levinson et al., 2011; Lindner et al., 2004; Tae-Hyun et al., 2006; Ziegler et al., 2014). Lindner et al. (2004) first introduced the concept of localization and map information based TLR. The system in that paper consisted of three parts: detector, tracker, and classifier. That paper proposed three different methods for detection, including color-based, shape-based, and cascade classifier-based methods. Among these methods, the paper found that a fusion method integrating a color-based detector, differential GPS, and map information was superior. In particular, the paper showed that when the system has 1 m of positioning accuracy and 1° of heading uncertainty, false positives can be reduced by five times. Tae-Hyun et al. (2006) introduced a guidance module for the TLR system. The guidance module provides the position of the ego vehicle, intersection information, and TL location with several points of attribute data. The map information is utilized for three purposes: as a task trigger for algorithm operation, to limit the search area in an image, and to estimate the size of a TL. Fairfield and Urmson (2011) introduced in detail several advantages when using extended digital map information for TLR: restricted scan areas, the composition of a robust classifier, and enhanced tracking performance. That paper highlighted a method for TL mapping by using image-to-image association and least squares triangulation. As a result, a huge TL map containing over 4000 sites was added to Google maps, and TLR was successfully conducted with a TL map and LIDAR (LIght Detection And Ranging) based precise localization during both day and night. Levinson et al. (2011) proposed another method for TL mapping that sequentially applies tracking, back-projection, and triangulation. The distinguishable contribution here is probabilistically estimating the TL state using various factors such as sensor data and uncertainty along with the relationship of multiple TLs at intersections. John et al. (2014) proposed a method to limit scan area using GPS and map information. In particular, normal and low illumination conditions are categorized into two specific environments and TLR is conducted based on a saliency map and convolutional neural network. In Ziegler et al. (2014), there are two modes for TLR: an offline mode that registers the distinctive visual features around TLs in a database, and an online mode that compares and matches registered features with extracted features in real-time to determine the location of the TL. This method successfully recognized 155 TLs on the Bertha Benz Historical Route. Barnes et al. (2015) introduced a probabilistic framework based on integrating a TL map and localization uncertainty for TLR to enhance online detection performance. This approach generates scale-space candidate regions, and evaluates detection scores based on classification results and the prior distribution of 3D occurrences.