ISPRS Journal of Photogrammetry and Remote Sensing 125 (2017) 184–192

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

A Bayesian approach to traffic light detection and mapping

Siavash Hosseinyalamdary *, Alper Yilmaz

Photogrammetric Computer Vision Laboratory (PCVLab), The Ohio State University, 2070 Neil Avenue, Columbus, OH 43210, USA

ARTICLE INFO

Article history:
Received 13 May 2016
Received in revised form 22 September 2016
Accepted 10 January 2017
Available online 3 February 2017

Keywords:
Traffic light detection and mapping
Conic section geometry
Spatio-temporal consistency
Bayesian inference

ABSTRACT

Automatic traffic light detection and mapping is an open research problem. The traffic lights vary in color, shape, geolocation, activation pattern, and installation which complicate their automated detection. In addition, the image of the traffic lights may be noisy, overexposed, underexposed, or occluded. In order to address this problem, we propose a Bayesian inference framework to detect and map traffic lights. In addition to the spatio-temporal consistency constraint, traffic light characteristics such as color, shape and height is shown to further improve the accuracy of the proposed approach. The proposed approach has been evaluated on two benchmark datasets and has been shown to outperform earlier studies. The results show that the precision and recall rates for the KITTI benchmark are 95.78% and 92.95% respectively and the precision and recall rates for the LARA benchmark are 98.66% and 94.65%.

1. Introduction

Accurate traffic light detection and mapping is an important task for autonomous vehicles (Diaz et al., 2015; Jensen et al., 2016). An autonomous vehicle should be able to detect the traffic lights and take proper actions based on the signal of traffic lights. Despite the fact that the autonomous driving technology is emerging, the traffic light detection is still an open challenge.

There are a number of challenges to detecting a traffic light: typically, the lenses on a traffic light are not illuminated uniformly and the lens color changes from center to its border, traffic lights may be installed on a pole or suspended over the road, and road regulations may change from one state to another. In addition to detection, the traffic lights needs to be geolocalized for a number of reasons: autonomous vehicle should stop in an appropriate distance from the traffic lights, and multiple traffic lights on a road should be sorted to select the correct traffic light for the autonomous vehicle. In order to overcome the aforementioned problems and geolocate the traffic lights, we propose a Bayesian probabilistic framework.

Color has been the prominent property of the traffic lights in the heuristic approaches. Color only based detection, however, has limitations due to noisy data acquisition. The lenses on a traffic light are not standard and have different shades of colors, and over saturation becomes a problem when camera directly faces the traffic lights. The Red-Green-Blue (RGB) color space is not suitable for the traffic light detection since its channels are not independent. Other color spaces separate the luma (image intensity) and chroma (color) components and therefore, they are more robust to the lighting changes and shadows. Researcher has explored various color spaces such as normalized RGB (Diaz-Cabrera and Cerri, 2013; Diaz-Cabrera et al., 2012, 2015; Omachi and Omachi, 2009, 2010), Hue-Saturation-Value (HSV) (Jie et al., 2013; Tae-Hyun et al., 2006), YCbCr (Cai et al., 2012), YUV (Shadeed et al., 2003), and CIELab (John et al., 2014; Sooksatra and Kondo, 2014). Some other researchers suggested to use multiple exposures and improve the illumination in the images (Jang et al., 2014).

There are a few researchers who applied brightness of the traffic light to detect them. The connected pixels can be matched with a resizable template of the traffic light.

The traffic light has a circular shape. If the image plane and traffic light plane are parallel, the circular shape of the traffic light in the image remains a circle. In order to exploit this characteristic, authors of Caraffi et al. (2008) and Huang and Lee (2010) use the Hough transform based circle detection. In order to overcome the computational complexity of the Hough transform, some researchers suggest the fast radial symmetry transform to detect circular shape of the traffic lights (Sooksatra and Kondo, 2014). Moreover, authors of de Charette and Nashashibi (2009a) detect the circular bright spots and apply adaptive template matching to find the traffic lights. The assumption that the traffic light fixture plane and image plane are parallel, is not always correct and the traffic light can have ellipse shape.

In addition to circular lens shape, the box-shaped fixture of the traffic light has also been explored. Unlike the traffic light lens, the

* Corresponding author.
E-mail address: hosseinyalamdary.1@osu.edu (S. Hosseinyalamdary).

http://dx.doi.org/10.1016/j.isprsjprs.2017.01.008
0924-2716/

traffic light fixture does not have primitive shape. Hence, template matching became a popular approach to detect the traffic light fixture (de Charette and Nashashibi, 2009a,b; Trehard et al., 2014; Wang et al., 2011). In addition, the AdaBoost classifier different classifiers has been also applied to detect the traffic light fixture (Gong et al., 2010; Kim et al., 2013).

The prior knowledge of traffic lights are essential for some traffic light detection algorithms. Since the traffic lights are static objects, they are geolocalized and stored in geospatial database. If intrinsic and extrinsic parameters of camera are known and the pose of platform is observed, the position of the traffic lights is projected into the image space and applied to initialize the traffic light detection algorithms (Barnes et al., 2015; Fairfield and Urmson, 2011; John et al., 2014; Levinson et al., 2011).

There are a number of approaches apply learning algorithms to detect the traffic lights. Convolutional Neural Network (CNN) has been applied to generate the saliency map and detect the traffic lights (John et al., 2014, 2015). In addition, it has been shown the Aggregated Channel Features (ACF) approach has supperior performance over the heuristic models (Jensen et al., 2016; Philipsen et al., 2015).

There are a number of shortcomings in the previous approaches: The geometry of traffic light lenses is neglected or poorly applied; Various features are not integrated in a statistical framework; Since the properties of traffic lights significantly vary in each state, evaluating the results on one dataset is not sufficient. Based on our knowledge, we are the first who has used conic sections to detect and localize the traffic lights. In addition, we utilize the Bayesian framework to combine several features and enforce spatiotemporal consistency. We evaluate our results using two benchmarks and compare them with the other approaches.

2. Methodology

In our approach, traffic light detection is formulated as a binary labeling problem and the traffic light characteristics such as color, shape, and height are used as observations for the traffic light detection. To ensure the detection is coherent in space and time, we additionally introduce spatio-temporal constraints.

2.1. Binary labeling

Suppose an image I_t is taken at time t and $\mathbf{x}_i = [u_i, v_i]^\top$ is one of its pixels. State $\omega_t(\mathbf{x}_i)$ indicates whether $\mathbf{x}_i$ belongs to the traffic light. In other words, $\omega_t(\mathbf{x}_i)$ is 1 if $\mathbf{x}_i$ belongs to a traffic light and it is 0 otherwise.

The observation vector $\mathbf{Z}_t(\mathbf{x}_i)$ is a vector of cues such as color, shape, and height of the traffic lights at time t. The best estimate of the traffic lights is calculated when all observations up to this time, $\mathbf{Z}_{1:t}(\mathbf{x}_i)$, are used. Therefore, probability of a pixel belongs to the traffic lights is given by $P(\omega_t(\mathbf{x}_i) = 1|\mathbf{Z}_{1:t}(\mathbf{x}_i))$. If probability of a pixel is sufficiently high, we label the pixel as a pixel of traffic light, such that:

$$\omega_t(\mathbf{x}_i) = \begin{cases} 1 & P(\omega_t(\mathbf{x}_i) = 1|\mathbf{Z}_{1:t}(\mathbf{x}_i)) > Th \\ 0 & \text{otherwise}, \end{cases} \tag{1}$$

where Th is empirically selected based on the precision and recall rate and is described later in the experiment section. The posterior probability of the labels for pixels of an image are estimated by:

$$P(\omega_t(\mathbf{x}_{1:n})|\mathbf{Z}_{1:t}(\mathbf{x}_{1:n})) = \frac{P(\mathbf{Z}_t(\mathbf{x}_{1:n})|\omega_t(\mathbf{x}_{1:n}))\,P(\omega_t(\mathbf{x}_{1:n})|\mathbf{Z}_{1:t-1}(\mathbf{x}_{1:n}))}{\int P(\mathbf{Z}_t(\mathbf{x}_{1:n})|\omega_t(\mathbf{x}_{1:n}))P(\omega_t(\mathbf{x}_{1:n})|\mathbf{Z}_{1:t-1}(\mathbf{x}_{1:n}))d\omega_t(\mathbf{x}_{1:n})}, \tag{2}$$

where n indicates to the number of pixels in the image I_t and $\mathbf{x}_{1:n}$ represents pixels of the image. In this equation, $P(\mathbf{Z}_t(\mathbf{x}_{1:n})|\omega_t(\mathbf{x}_{1:n}))$ is the likelihood term and relates to the observation vector and the labels at the current time. The $P(\omega_t(\mathbf{x}_{1:n})|\mathbf{Z}_{1:t-1}(\mathbf{x}_{1:n}))$ term is the prior term and it relates the labels to the previous observations. The denominator is a normalization term to enforce probability range $[0, 1]$.

We estimate the likelihood term of (2) by computing joint probability across all pixels:

$$P(\mathbf{Z}_t(\mathbf{x}_{1:n})|\omega_t(\mathbf{x}_{1:n})) = \prod_{i=1}^{n} P(\mathbf{Z}_t(\mathbf{x}_i)|\omega_t(\mathbf{x}_i))P(\omega_t(\mathbf{x}_i)|\omega_t(\mathbf{x}_{i\notin 1:n})), \tag{3}$$

where $\mathbf{x}_{i\notin 1:n}$ represents all pixels of an image expect pixel $\mathbf{x}_i$ and the term $P(\omega_t(\mathbf{x}_i)|\omega_t(\mathbf{x}_{i\notin 1:n}))$ indicates the probability of the pixel $\mathbf{x}_i$ condition to the probability of the other pixels. If the spatial correlation between pixels are neglected, then $P(\omega_t(\mathbf{x}_i)|\omega_t(\mathbf{x}_{i\notin 1:n})) = P(\omega_t(\mathbf{x}_i))$.

Assuming Markovian condition, the labels at the current time depend only on the previous time and therefore $\omega_t(\mathbf{x}_{1:n})$ and $\omega_{1:t-2}(\mathbf{x}_{1:n})$ are independent given $\omega_{1:t-1}(\mathbf{x}_{1:n})$. Furthermore, the prior term of (1) is calculated from marginalization of $P(\omega_{1:t}(\mathbf{x}_{1:n})|\mathbf{Z}_{1:t-1}(\mathbf{x}_{1:n}))$ over $\omega_{1:t-1}(\mathbf{x}_{1:n})$:

$$P(\omega_t(\mathbf{x}_{1:n})|\mathbf{Z}_{1:t-1}(\mathbf{x}_{1:n})) = \Sigma_{i=0}^{1} P(\omega_t(\mathbf{x}_{1:n})|\omega_{t-1}(\mathbf{x}_{1:n}) = i)P(\omega_{t-1}(\mathbf{x}_{1:n}) = i|\mathbf{Z}_{1:t-1}(\mathbf{x}_{1:n})). \tag{4}$$

In this equation, $P(\omega_{t-1}(\mathbf{x}_{1:n}) = i|\mathbf{Z}_{1:t-1})$ term is the posterior estimation of the state in the previous time and the $P(\omega_t(\mathbf{x}_{1:n})|\omega_{t-1}(\mathbf{x}_{1:n}) = i)$ is transition term that predicts the state in the current time based on its estimate in the previous time.

2.2. Spatial coherency

Generally speaking, it is most likely to have the neighboring pixels with similar color belonging to the same object. Therefore, the probability that these pixels have the same label can be computed by:

$$P(\omega_t(\mathbf{x}_i)|\omega_t(\mathbf{x}_j)) = \lambda(1 - \delta(\omega_t(\mathbf{x}_i) - \omega_t(\mathbf{x}_j))) \times \exp\left(-\frac{||I(\mathbf{x}_i) - I(\mathbf{x}_j)||^2}{2\beta}\right), \tag{5}$$

where λ is a constant value, δ is delta-kroneker function, and $\lambda(1 - \delta(\omega_t(\mathbf{x}_i) - \omega_t(\mathbf{x}_j))$ enforces the probability to be between zero and one. In addition, β is the average of color variations and is estimated:

$$\beta = \frac{1}{n}\sum_{i=1}^{n}\sum_{j}||I(\mathbf{x}_i) - I(\mathbf{x}_j)||^2. \tag{6}$$

2.3. Temporal constraint

In Eq. (4), $P(\omega_{t-1}(\mathbf{x}_{1:n})|\mathbf{Z}_{1:t-1}(\mathbf{x}_{1:n}))$ is the posterior estimation of the state probability in the previous time. In addition, $P(\omega_t(\mathbf{x}_{1:n})|\mathbf{Z}_{1:t-1}(\mathbf{x}_{1:n}))$ is the predicted state probability at the current time since the observations in the current time, $\mathbf{Z}_t$, are not used. $P(\omega_t(\mathbf{x}_{1:n})|\omega_{t-1}(\mathbf{x}_{1:n}))$ is applied to predict the state probability based on the posterior estimation of this probability in the previous time.

Let's assume that the relationship between the labels of current and previous times is linear, such that:

$$\omega_t(\mathbf{x}_i) = \mu_p + \Psi\omega_{t-1}(\mathbf{x}_i) + \epsilon_p, \tag{7}$$

where μ_p is the constant change of labels from previous epoch to the current time, Ψ is the vector that contains the coefficients of the linear function between the current and previous time and ϵ_p is noise. Conjecturing that noise in the transition is normally distributed, $P(\omega_t(\mathbf{x}_{1:n})|\omega_{t-1}(\mathbf{x}_{1:n}))$ is computed by: