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

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 $\omega_{1:t-1}(\mathbf{x}_{1:n})$:

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

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:

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

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:

$$P(\omega_t(\mathbf{x}_i)|\omega_{t-1}(\mathbf{x}_i)) \sim \mathcal{N}_\omega(\mu_p + \Psi\omega_{t-1}(\mathbf{x}_i), \sigma_p^2), \tag{8}$$

where $\mathcal{N}(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2.

In the case when the platform is stationary, two images are the same and $P(\omega_t(\mathbf{x}_i)|\omega_{t-1}(\mathbf{x}_i))$ is normally distributed when $\mu_p = 0, \Psi = 1$. Therefore, changes in the labels may occur due to noise. In contrast, when the platform is in motion, the dynamics of the platform will affect two coefficients, μ and ψ. Since the traffic light which is seen from an oblique view is planar, the relationship between the real world coordinates of the traffic light and their projection into the image space becomes a homography transformation, $\mathtt{H}$. If the projection of a 3D point, $\mathbf{X}_i$, of the traffic light to the current and the previous images are $\mathbf{x}_i = \mathtt{H}_{t-1}\mathbf{X}_i$ and $\mathbf{x}'_i = \mathtt{H}_t\mathbf{X}_i$, then it can be shown that $\mathbf{x}'_i = \mathtt{H}_t\mathtt{H}_{t-1}^{-1}\mathbf{x}_i$, such that it follows also a normal distribution:

$$P(\omega_t(\mathbf{x}'_i)|\omega_{t-1}(\mathbf{x}_i)) \sim \mathcal{N}_\omega(\omega_{t-1}(\mathtt{H}_t\mathtt{H}_{t-1}^{-1}\mathbf{x}_i), \sigma_p^2). \tag{9}$$

For the first time where the previous label estimation is not available, the probability of the labels are assumed to follow a binomial distribution:

$$P(\omega_1(\mathbf{x}_i)) = \begin{cases} \kappa & \omega_1(\mathbf{x}_i) = 1 \\ 1-\kappa & \omega_1(\mathbf{x}_i) = 0, \end{cases} \tag{10}$$

where κ is a constant learned from existing data and $P(\omega_1(\mathbf{x}_i) = 0) + P(\omega_1(\mathbf{x}_i) = 1) = 1$.

2.4. Evidence

The traffic light has an active lens that can be either red, yellow or green, and has a circular shape. The traffic light may be suspended or installed on a pole and therefore, its height follows the installation standards. The activation pattern of the traffic light lenses also follows regulations which may be used to detect the traffic lights. Last but not least, the geolocation of the traffic lights may be retrieved from GIS maps and applied to detect the traffic lights.

Let's define an observation vector, $\mathbf{Z}_t(\mathbf{x}_i) = \{z_c, z_s, z_h, z_i, z_g\}$, where the vector includes the color, shape, height, inactive lenses pattern and GIS cues of the traffic light, respectively. The observation vector is given by:

$$\begin{aligned} &P(\mathbf{Z}_t(\mathbf{x}_i)|\omega_t(\mathbf{x}_i)) = P(z_i|z_h, \omega_t(\mathbf{x}_i))P(z_h|z_s, \omega_t(\mathbf{x}_i)) \\ &P(z_s|z_c, \omega_t(\mathbf{x}_i))P(z_c|\omega_t(\mathbf{x}_i))P(z_g|\omega_t(\mathbf{x}_i)). \end{aligned} \tag{11}$$

The color characteristic generally discriminates the traffic lights from the other objects. However, observed color may change due to the different illumination conditions and camera response. In order to utilize the color feature, let's define a hidden variable, h_k, that represents the red, yellow, and green colors of the traffic light for $k \in \{1, 2, 3\}$. These colors have been shown to be normally distributed with mean μ_k and variance $\sigma_k^2, \mathcal{N}_{\omega,k}(\mu_k, \sigma_k^2)$ and shown in Fig. 1a. Therefore, the probability of the color cue is the mixture of Gaussians. In addition, the color cue for the background can be assumed as normal distribution, $\mathcal{N}_\omega(\mu_0, \sigma_0^2)$, such that:

$$P(z_c|\omega_t(\mathbf{x}_i)) = \begin{cases} \Sigma_{k=1}^3 w_k \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(h-\mu_k)^2}{2\sigma_k^2}\right) & \omega_1(\mathbf{x}_i) = 1 \\ \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(h-\mu_0)^2}{2\sigma_0^2}\right) & \omega_1(\mathbf{x}_i) = 0, \end{cases} \tag{12}$$

where w_k is the normalization coefficient and it guarantees that the probability is between zero and one.

The traffic lights are circular and can transform to an ellipse under perspective geometry. Using ellipses is important to estimate the projective transformation, and consequently, the depth of the traffic light. In the case of using circles, this geometric constraint will be violated. Hence, an object segmented based on its color is a traffic light if its shape is an ellipse. Ellipse is a conic section and can be represented by a 3×3 symmetric matrix, $\mathtt{C}$. If the pixel $\mathbf{x}_i$ belongs to the traffic light, it should reside inside the ellipse, $\bar{\mathbf{x}}_i^\top \mathtt{C} \bar{\mathbf{x}}_i \leqslant 0$, where $\bar{\mathbf{x}}_i = [\mathbf{x}_i; 1]$ is in homogeneous coordinates representation of the pixel $\mathbf{x}_i$. The probability of this inequality can be represented by a Chi distribution with one degree of freedom, such that:

$$P(z_s|\omega_t(\mathbf{x}_i)) = \begin{cases} \frac{\sqrt{2}e^{-\frac{\bar{\mathbf{x}}_i^\top \mathtt{C} \bar{\mathbf{x}}_i}{2}}}{\Gamma\left(\frac{1}{2}\right)} & \omega_1(\mathbf{x}_i) = 1 \\ \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{\bar{\mathbf{x}}_i^\top \mathtt{C} \bar{\mathbf{x}}_i}{2\sigma_0^2}\right) & \omega_1(\mathbf{x}_i) = 0, \end{cases} \tag{13}$$

where Γ is the gamma function in Chi distribution. Fig. 1c illustrates the color based probability of pixels inside the traffic light. The pixels that are inside representing the traffic light shape have higher probability than the ones on the borders and outside.

The homography transformation can estimate the position of the traffic light with respect to the camera coordinate system, $\mathbf{X} = \mathtt{H}^{-1}\mathbf{x}$, by back-projecting the center of candidate traffic light into the object space, which also provides a means to estimate the height of the traffic light. If the camera is aligned with vertical direction and the height of the camera from the ground is known, the height of the traffic light can be estimated. Let's assume that a traffic light can be installed in K different heights based on the traffic light installation standards. The height cue is a mixture of Gaussians (as shown in Fig. 1b), such that:

$$P(z_h|z_s, \omega_t(\mathbf{x}_i)) = \begin{cases} \Sigma_{k=1}^K w_k \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(h-\mu_k)^2}{2\sigma_k^2}\right) & \omega_1(\mathbf{x}_i) = 1 \\ \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(h-\mu_0)^2}{2\sigma_0^2}\right) & \omega_1(\mathbf{x}_i) = 0. \end{cases} \tag{14}$$

Considering that when one of the lenses is active, other lenses are inactive. Since the traffic light dimensions are standard, the position of the other lenses can be calculated. If one of the traffic light lenses are active, the other lenses should be dark. Since the gray value of the dark lens is non-negative, it follows the half-normal distribution, such that:

$$P(z_i|z_h, \omega_t(\mathbf{x}_i)) = \begin{cases} \frac{2\theta}{\pi} \exp\left(-\frac{1}{I}(\mathbf{x}_i^2)\theta^2\pi\right) & \omega_1(\mathbf{x}_i) = 1 \\ \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(I(\mathbf{x}'_i)-\mu_i)^2}{2\sigma_i^2}\right) & \omega_1(\mathbf{x}_i) = 0, \end{cases} \tag{15}$$

where $I(\mathbf{x}'_i)$ is the gray value of the inactive lenses and $\mu = \frac{1}{\theta}$. Fig. 1d illustrates the half-distribution of the green lens when it is inactive. The background is also modeled as normal distribution in (15).

In some countries, the traffic lights may be installed not only vertically, but also horizontally. Therefore, the inactive lenses of the traffic lights may be searched above and below the active lens or left and right sides of it. In addition, the activation pattern of the lights differ and yellow light can be activated before or after red light. It also can be activated with the red or green lights.

In the case when GIS maps are available, position of the traffic lights can be retrieved and projected into the image space. Unfortunately, the accurate map production is not trivial and there are only a few GIS maps that contain accurate positions of the traffic lights. For instance, OpenStreetMap is a public GIS, and contains inaccurate 2D geolocation of the traffic lights. Since the users provide the information in OpenStreetMap, its accuracy and completeness is not sufficient for direct projection into the image. In addition, the imperfect navigation solution leads to uncertainty in position of the projected traffic light. Therefore, probability that the pixel $\mathbf{x}_i$ belongs to the traffic light depends on the accuracy of

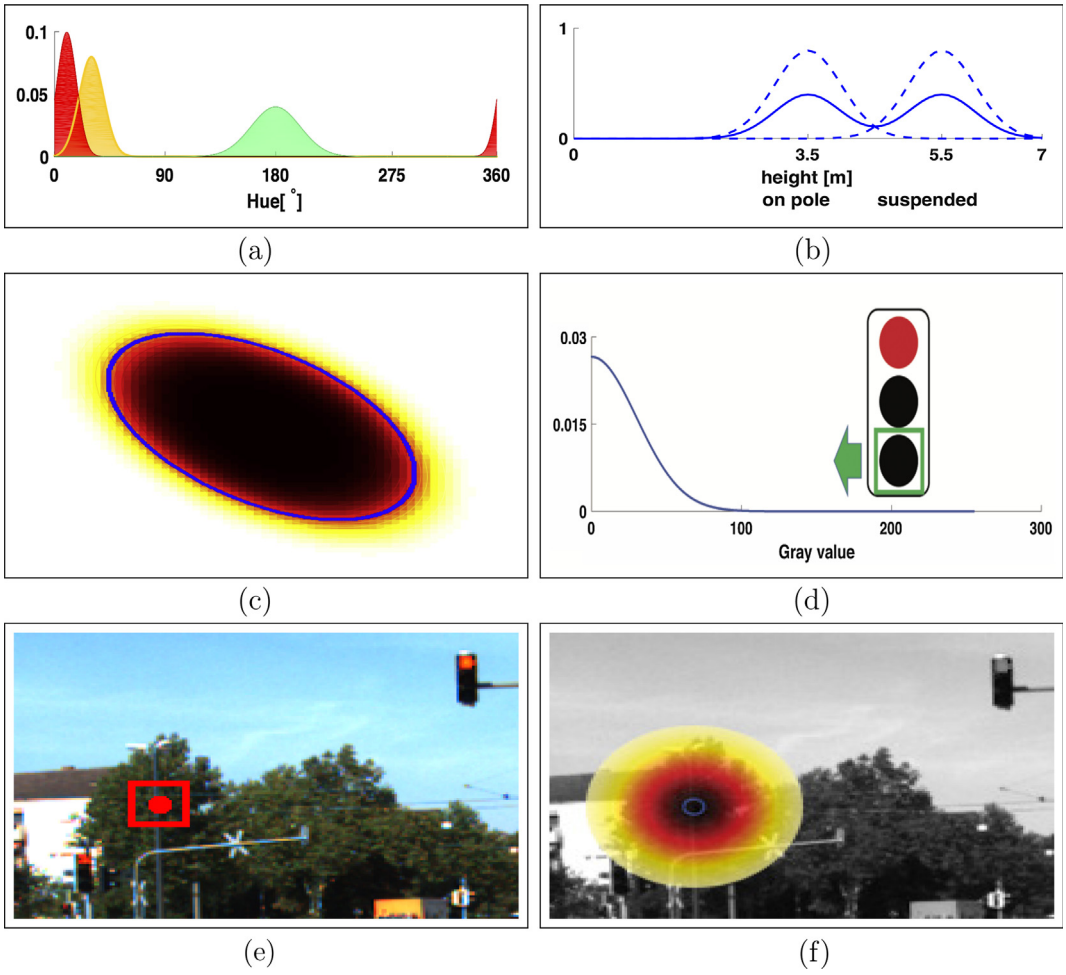


Fig. 1. Observations of the traffic light (a) The color characteristics of the traffic lights are observed in HSV color space and these colors are represented as mixture of Gaussians in hue components. The red color distribution is continuous since the hue component is circular and it is in the range $[0\ 2\pi]$. (b) The mixture of two Gaussians (solid line) is used to model height: the one for the traffic lights on the pole and the one for the suspended traffic lights on the road (dashed lines). (c) The probability of a pixel belonging to the ellipse is 1 within the ellipse, it is zero outside the ellipse and it is modeled by Chi distribution. (d) When red signal is active and the green lens should be inactive and dark. Therefore, the pixels of green lens follow half-normal distribution. (e) The traffic light is detected, its position is retrieved from OSM, and it is projected to the image space. A rectangle is added to improve the visualization. (f) The probability of the labels based on GIS cue is shown by color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

projected position of the traffic light in the GIS maps. This probability follows a bivariate normal distribution, such that:

$$P(z_g|\omega_t(\mathbf{x}_i)) = \begin{cases} \frac{1}{\sqrt{2\pi\Sigma_g^2}} \exp\left(-\frac{(\mathbf{x}_i-\tilde{\boldsymbol{\mu}}_g)^2}{2\Sigma_g^2}\right) & \omega_1(\mathbf{x}_i)=1 \\ \frac{1}{\sqrt{2\pi\Sigma_0^2}} \exp\left(-\frac{(\mathbf{x}_i-\tilde{\boldsymbol{\mu}}_0)^2}{2\Sigma_0^2}\right) & \omega_1(\mathbf{x}_i)=0, \end{cases} \tag{16}$$

where $\tilde{\boldsymbol{\mu}}_g$ is the projected traffic light from the GIS map to image space. The traffic light shown with a red rectangle in Fig. 1e is retrieved from GIS maps and projected to the image space. The probability of the state based on GIS cue is shown by color in Fig. 1f.

2.5. Learning

The statistical models, relates the labels with observations, are assumed to be known in the previous sections. However, Some parameters of these statistical models such as mean and variance may not be known beforehand. The unknown parameters can be learned in an Expectation Maximization framework. For simplicity, let's assume that $\mathcal{N}(\mu_0^{[j-1]}, \Sigma_0^{[j-1]})$ is the background model and $\mathcal{N}(\mu_1^{[j-1]}, \Sigma_1^{[j-1]})$ is the traffic light model. The labels have binomial distribution, $bin(\kappa)$. Therefore, the unknown parameters are $\boldsymbol{\mu}_0, \Sigma_0, \boldsymbol{\mu}_1, \Sigma_1$, and κ that should be learned in this process. The labels are estimated in the E-step:

$$q^{[j]}(\omega_t(\mathbf{x}_i)) = \begin{cases} \frac{\kappa\mathcal{N}(\mu_1^{[j-1]},\Sigma_1^{[j-1]})}{\kappa\mathcal{N}(\mu_0^{[j-1]},\Sigma_0^{[j-1]})+(1-\kappa)\mathcal{N}(\mu_0^{[j-1]},\Sigma_0^{[j-1]})} & \omega_1(\mathbf{x}_i)=1 \\ \frac{(1-\kappa)\mathcal{N}(\mu_0^{[j-1]},\Sigma_0^{[j-1]})}{\kappa\mathcal{N}(\mu_0^{[j-1]},\Sigma_0^{[j-1]})+(1-\kappa)\mathcal{N}(\mu_0^{[j-1]},\Sigma_0^{[j-1]})} & \omega_1(\mathbf{x}_i)=0, \end{cases} \tag{17}$$

where q is $P(\omega_t(\mathbf{x}_i)|\mathbf{Z}_t(\mathbf{x}_i), \mu_0^{[j-1]}, \Sigma_0^{[j-1]})$. The parameters of statistical models are estimated in M-step based on the estimated labels, $q^{[j]}(\omega_t(\mathbf{x}_i))$, such that:

$$\begin{cases} \hat{\mu}_1^{[j]}, \hat{\Sigma}_1^{[j]} = \underset{\mu_0,\Sigma_0}{\arg\max}(\Sigma_{i=1}^n q^{[j]}(\omega_t(\mathbf{x}_i)) \log[\kappa\mathcal{N}(\mu_1^{[j-1]}, \Sigma_1^{[j-1]})]) & \omega_1(\mathbf{x}_i)=1 \\ \hat{\mu}_0^{[j]}, \hat{\Sigma}_0^{[j]} = \underset{\mu_0,\Sigma_0}{\arg\max}(\Sigma_{i=1}^n q^{[j]}(\omega_t(\mathbf{x}_i)) \log[(1-\kappa)\mathcal{N}(\mu_0^{[j-1]}, \Sigma_0^{[j-1]})]) & \omega_1(\mathbf{x}_i)=0, \end{cases} \tag{18}$$

The superscript j is iteration of the expectation maximization in the E-step and M-step.

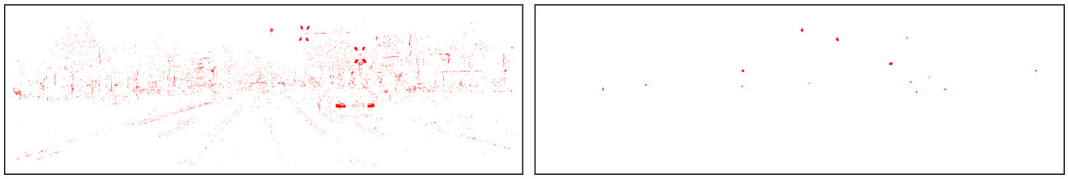


Fig. 2. Search space reduction: (a) red regions are selected using a red color mask; (b) non-ellipse-shaped regions are removed and the search space is significantly reduced. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
The initial parameters of the models are applied in this paper. These parameters are updated in the learning process.

	$P(\omega)$	$P(Z_c\|\omega)$	$P(Z_g\|\omega)$	$P(Z_h\|\omega)$	$P(Z_i\|\omega)$
Distribution	Binomial	MOG	Gaussian	MOG	Gaussian
$P(\omega=1)$	$\kappa=10^{-4}$	$N_1(10,8)$ $N_2(30,10)$ $N_3(180,20)$	$N_1(\tilde{\boldsymbol{\mu}},30)$	$N_1(3.5,0.5)$ $N_1(5.5,0.5)$	$N_1(0,30)$
$P(\omega=0)$	$1-\kappa$	$N_0(180,90)$	$N_0(\tilde{\boldsymbol{\mu}},100)$	$N_0(0,20)$	$N_0(125,70)$

3. Implementation

The proposed approach estimates the probability of each pixel belonging to the traffic light. In order to speed up our proposed approach, the search space is reduced and many pixels that are less likely to be traffic light are removed before we apply the proposed approach to the remained pixels. We also discuss the initial values of the statical models in this section.

3.1. Reduction of the search space

Ideally, all pixels of the image should be contributed in the proposed Bayesian framework. However, it is computationally expensive, hence, we reduce the search space for the traffic lights detection. In order to remove the pixels that are less likely to belong to the traffic light, color masks are utilized to remove the objects that do not have red, yellow or green colors. In addition, objects with low saturation and value are either too bright or too dark and are removed. For remaining pixels, the ellipses are fitted to connecting pixel regions and the ones with high fitting error are removed. The remaining connected pixel regions, also called objects, are marked as potential traffic lights and we apply the proposed Bayesian framework to these objects to detect traffic lights. The red color mask is applied to an image in Fig. 2a and the ellipse fitting is used to reduce the search space as shown in Fig. 2b.

3.2. Initial parameters

The parameters of statistical models are learned in the learning process as previously explained. However, the initial value of these parameters are required in (17) and (18). Table 1 demostrates the initial value of these parameters.

In the GIS cue, $\tilde{\boldsymbol{\mu}}$ is the projected traffic lights from the database into the image space, and therefore, it is known and does not need to be learned in learning process. The binomial constant, κ, is selected in the way that it represents a 5×5 traffic light in a 640×480 image. The red, yellow and green normal distributions follow the histograms in Levinson et al. (2011). The distributions should be chosen in the way that integration of each distribution function becomes one. For instance, if the color range is between 0 and 255, it should be ensured that the probability of the distribution function that goes beyond the color range is negligible and its integration is one within this range.

4. Experiments

We have applied our proposed Bayesian framework to two publicly available benchmark datasets: Karlsruhe Institute of Technology (KITTI) and La Route Automatise (LARA) benchmarks. In KITTI dataset, multiple calibrated and synchronized sensors are mounted on a platform and the data is collected in Karlsruhe, Germany. The ground truth for the traffic lights are not given and we have manually annotated the traffic lights in the images. The LARA benchmark is specific for traffic light detection and the ground truth is given. This dataset is collected from downtown of Paris, France, utilizing an uncalibrated camera on the platform. Since navigation solution is not available for the LARA benchmark, the GIS cue cannot be used for in the proposed traffic light detection approach. In addition, height of the camera is not given in the LARA benchmark and an uncertainty is introduced in the height cue. Traffic lights in these benchmarks have different properties such as size and color. For instance, the traffic lights are smaller and lower in the LARA benchmark. On the other hand, the traffic lights in the KITTI dataset can be installed on poles or suspended.

The camera applied in KITTI benchmark has 1027×768 resolution and 4 mm focal length (Geiger et al., 2013) and the one in LARA benchmark has 640×480 pixels resolution and 12 mm focal length. The KITTI dataset contains a total of 404 red lights, 10 yellow lights, and 26 green lights and the LARA dataset contains 5280 red lights, 58 yellow lights, 3381 green lights. Since the LARA benchmark is lengthy with many stationary frames, we focus on the part that vehicle drives in downtown of Paris. This part has more than 1800 frames and it contains 2486 traffic lights.

In order to quantitatively evaluate our proposed approach, we apply the PASCAL criterion, utilized in context of object detection (Yilmaz et al., 2006). The detected traffic lights are defined as bounding boxes. The PASCAL criterion labels an object as a correctly detected traffic light if the intersection of the bounding box and the ground truth is more than half on the union of them. Otherwise, the detected traffic lights are labeled as false positives. The true negatives are not relevant and we calculate the true positives (TP), false positives (FP), and false negatives (FN).

In this paper, the proposed approach is evaluated and compared to the previous work based on the precision and recall criteria. The precision is the ratio of true positives to the positive outcomes:

$$Precision = \frac{TP}{TP+FP}. \tag{19}$$

The recall criterion is the ratio of true positives to the number of traffic lights, $TP + FN$, such that:

$$Recall = \frac{TP}{TP + FN}. \quad (20)$$

If the probability a pixel belongs to the traffic light is higher than a threshold in (1), we label the pixel as traffic light. In order to find the optimum threshold, precision and recall criteria are plotted based on different thresholds and the one with maximum precision and recall rate is chosen. Moreover, we apply F-score criterion to compare the results. In F1-score, the precision and recall rates are evenly weighted, such that:

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \quad (21)$$

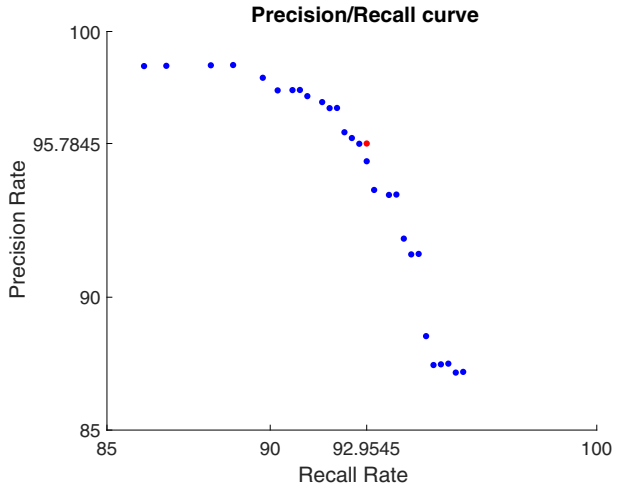


Fig. 3. Precision/recall curve for the KITTI dataset. The best precision/recall rate balance occurs at $P(\omega_t) > 88.4\%$. Number of the traffic lights is 440.

4.1. KITTI dataset performance

The precision and recall rates are plotted with respect to different thresholds in (1) and shown in Fig. 3.

The maximum precision and recall rates are estimated such that it distance to the perfect solution, with 100% precision and recall rates, becomes minimum. The precision and recall rates is maximum when the threshold is 88.4%. The precision and recall rates are 95.8% and 93.0%.

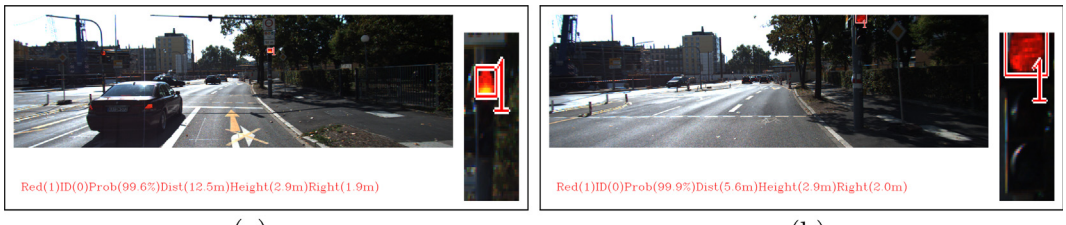
It should be verified wether the active color of the traffic light is correctly recognized. The yellow traffic light has been active once and it was not sufficiently observed. Therefore, we merged the yellow and red traffic lights. The results of the detected traffic lights based on their color are tabulated in Table 2. The confusion matrix shows that all of the traffic light signals are correctly recognized and there is no intra-class confusion.The traditional traffic lights are assembled by a light source and a color filter. Therefore, the color of lens varies if the light energy is not uniformly distributed on the lens. One instance of this situation has been shown in Fig. 4a. A part of the lens which is closer to light source is changed to yellow, although the traffic light signal is red. Therefore, a weighted voting scheme has been applied to recognize the color of every pixel that belongs to the traffic light. In addition, partial occlusion may occur in the traffic lights. In Fig. 4b, the partially occluded traffic light is still detectable since the temporal constraints has been utilized.

In Fig. 5, the platform has been waiting behind the red traffic light, the traffic light signal converts to yellow and consequently, green. The traffic light lens activation pattern may be different in different countries. In addition, the red and yellow lenses are simultaneously active and the proposed algorithm can correctly detect the active lenses of the traffic light. Since the red traffic light size is not precisely estimated, the distance of the red lens from camera has been incorrectly computed.

In Fig. 6, multiple traffic lights have been utilized to regulate the traffic in each lane. The traffic lights have been correctly detected and their location have been estimated. Using the estimated geolocation of the traffic lights, the platform can choose the one that is corresponding to its lane.

Table 2
The confusion matrix; It shows that the traffic light signals have been correctly detected.

			System Red 380	Classification Green 21
Ground Truth	Red Green	404 26	380 0	0 21
Recall rate			94%	81%



(a) (b)

Fig. 4. The traffic light detection for the KITTI dataset; (a) The lower part of the red signal has more illumination and it is transformed to yellow. Therefore, the color of the signal should be evaluated for every pixel and the choice of the color should be estimated in a voting scheme. (b) The traffic light is partially occluded, but the algorithm can detect the traffic light since it imposes the temporal consistency constraint. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 5. A scenario that the traffic light signal changes from (a) red to (b) yellow and consequently, (c) green in KITTI dataset. (d) There is a false positive that has low probability (61%) and will be removed using threshold in (1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

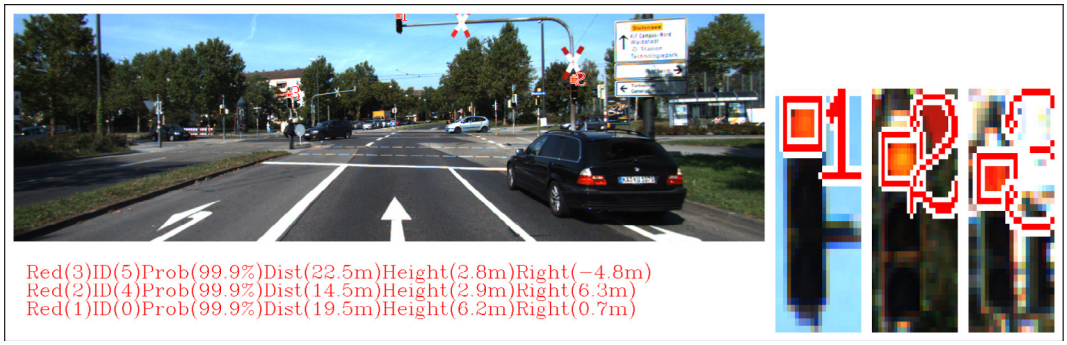


Fig. 6. Multiple traffic lights in KITTI dataset. The estimated geolocation of the traffic lights can be applied to find the traffic lights that correspond to the each lane.

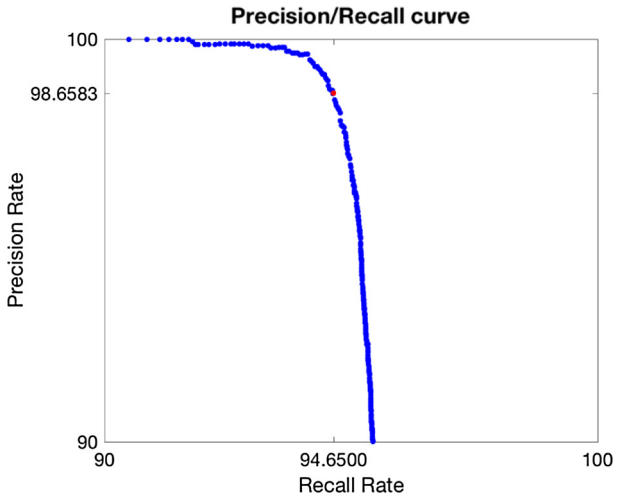


Fig. 7. Precision/recall curve for LARA dataset. The best precision and recall rate balance happens at $P(\omega_t) > 93.9\%$. Number of the traffic lights is 2486.

4.2. LARA dataset performance

The LARA benchmark has also been utilized to assess the proposed approach. The Precision/Recall curve has been plotted for the LARA dataset and it is presented in Fig. 7. The best precision and recall balance is reached where the threshold in (1) is 93.8%. By choosing this threshold, the precision and recall rates are 98.7% and 94.7%. The precision and recall rates were compared with the previous work on this dataset and the results are given in Table 3. The precision rate of the proposed approach is the highest among the previous approaches. Although the recall rate is not the highest among other approaches, it is comparable and the respective precision is the highest. Since the true negative is not given in the ground truth, the accuracy criterion cannot be estimated for these two approaches. Comparing F1-score of the proposed approach with the previous work shows that the proposed approach (96.61%.) and (de Charette and Nashashibi, 2009a) (96.89%.) have the highest F1-score.

Results of the proposed approach applied to LARA benchmark, has been shown in Fig. 8. We observed that the farther traffic lights are more difficult to be detected since they are represented by fewer pixels. This 15 s sequence of LARA benchmark results shows that the traffic lights can be detected as far as 40 meters and they are accurately tracked within the sequence.

Table 3
Comparison of the traffic light detection algorithms using LARA benchmark; Our proposed approach has the highest precision rate and the recall rate is hight, too.

	de Charette and Nashashibi (2009a)	de Charette and Nashashibi (2009b)	Haltakov et al. (2015)	Siogkas et al. (2012)	Wang et al. (2011)	Ours
Precision rate	95.38%	84.5%	72.83%	61.22%	96.95%	**98.66**%
Recall rate	**98.41**%	53.5%	80.13%	93.75%	94.4%	94.65%
F1 score	**96.89**%	65.52%	76.30%	74.07%	95.66%	**96.61**%

Fig. 8. The traffic lights are detected in LARA dataset (a) from beginning, (b) after 1 s, (c) after 3 s, (d) after 6 s, (e) after 11 s, (f) after 14 s. The signals are correctly recognized and the position of the traffic lights is estimated with respect to the camera.

5. Conclusion

We have introduced a Bayesian statistical framework to detect the traffic lights and recognize their signal. In order to preserve the coherency in space and time, a spatio-temporal consistency condition is applied. Several characteristics of traffic lights such as color, shape, height, inactive lenses pattern, and GIS cues are used as observations. The color is modeled as mixture of Gaussians and a Chi distribution is utilized to model the shape cue. The height cue is also modeled as mixture of Gaussians since traffic lights

can be installed on the pole or suspended and have different heights. We model the inactive lenses pattern as half-normal distribution and GIS cue is represented by bivariate Gaussian distribution. The conic section geometry has been applied in the proposed approach to estimate the pose of the traffic lights with respect to the camera coordinate system. We have evaluated the results of the proposed traffic light detection using two benchmarks and results outperform the earlier traffic light detection approaches.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.isprsjprs.2017.01.008.

References

Barnes, D., Maddern, W., Posner, I., 2015. Exploiting 3D semantic scene priors for online traffic light interpretation. In: Intelligent Vehicles Symposium (IV). IEEE, Seoul, pp. 573–578. June 28–July 1.

Cai, Z., Li, Y., Gu, M., 2012. Real-time recognition system of traffic light in urban environment. In: Symposium on Computational Intelligence for Security and Defence Applications (CISDA). IEEE, Ottawa, ON, Canada, pp. 1–6. 11–13 July.

Caraffi, C., Cardarelli, E., Medici, P., Porta, P.P., Ghisio, G., Monchiero, G., 2008. An algorithm for italian de-restriction signs detection. In: Intelligent Vehicles Symposium. IEEE, Eindhoven, Netherlands, pp. 834–840. 4–6 June.

de Charette, R., Nashashibi, F., 2009a. Real time visual traffic lights recognition based on spot light detection and adaptive traffic lights templates. In: Intelligent Vehicles Symposium. IEEE, Dearborn, MI, USA, pp. 358–363. 8–11 June.

de Charette, R., Nashashibi, F., 2009b. Traffic light recognition using image processing compared to learning processes. In: International Conference on Intelligent Robots and Systems. IEEE, St. Louis, USA, pp. 333–338. 11–15 October.

Diaz-Cabrera, M., Cerri, P., 2013. Traffic light recognition during the night based on fuzzy logic clustering. In: 14th International Conference on Computer Aided Systems. Lect. Notes Comput. Sci., vol. 8112. Springer, Berlin Heidelberg, Las Palmas de Gran Canaria, Spain, pp. 93–100. 10–15 February.

Diaz-Cabrera, M., Cerri, P., Sanchez-Medina, J., 2012. Suspended traffic lights detection and distance estimation using color features. In: 15th International Conference on Intelligent Transportation Systems (ITSC). IEEE, Anchorage, AK, USA, pp. 1315–1320. 16–19 September.

Diaz-Cabrera, M., Cerri, P., Medici, P., 2015. Robust real-time traffic light detection and distance estimation using a single camera. Expert Syst. Appl. 42 (8), 3911–3923.

Diaz, M., Cerri, P., Pirlo, G., Ferrer, M.A., Impedovo, D., 2015. A survey on traffic light detection. In: New Trends in Image Analysis and Processing. Lect. Notes Comput. Sci., vol. 9281. Springer International Publishing, pp. 201–208.

Fairfield, N., Urmson, C., 2011. Traffic light mapping and detection. In: International Conference on Robotics and Automation (ICRA). IEEE, Shanghai, China, pp. 5421–5426. 9–13 May.

Geiger, A., Lenz, P., Stiller, C., Urtasun, R., 2013. Vision meets robotics: the kitti dataset. Int. J. Robot. Res. (IJRR) 32, 1231–1237.

Gong, J., Jiang, Y., Xiong, G., Guan, C., Tao, G., Chen, H., 2010. The recognition and tracking of traffic lights based on color segmentation and camshift for intelligent vehicles. In: Intelligent Vehicles Symposium. IEEE, San Diego, CA, USA, pp. 431–435. 21–24 June.

Haltakov, V., Mayr, J., Unger, C., Ilic, S., 2015. Semantic segmentation based traffic light detection at day and at night. The German Conference on Pattern Recognition (GCPR), vol. 9358. Springer International Publishing, Aachen, Germany, pp. 446–457. 7–10 October.

Huang, Y.S., Lee, Y.S., 2010. Detection and recognition of speed limit signs. In: International Computer Symposium (ICS). Tainan, Taiwan, pp. 107–112. 16–18 December.

Jang, C., Kim, C., Kim, D., Lee, M., Sunwoo, M., 2014. Multiple exposure images based traffic light recognition. In: Intelligent Vehicles Symposium (IVS). IEEE, Dearborn, MI, USA, pp. 1313–1318. 8–11 June.

Jensen, M.B., Philipsen, M.P., Møgelmose, A., Moeslund, T.B., Trivedi, M.M., 2016. Vision for looking at traffic lights: Issues, survey, and perspectives. IEEE Trans. Intell. Transp. Syst. PP (99), 1–16.

Jie, Y., Xiaomin, C., Pengfei, G., Zhonglong, X., 2013. A new traffic light detection and recognition algorithm for electronic travel aid. In: 4th International Conference on Intelligent Control and Information Processing (ICICIP). IEEE, Beijing, China, pp. 644–648. 9–11 June.

John, V., Yoneda, K., Qi, B., Liu, Z., Mita, S., 2014. Traffic light recognition in varying illumination using deep learning and saliency map. In: 17th International Conference on Intelligent Transportation Systems (ITSC). IEEE, Qingdao, China, pp. 2286–2291. 8–11 October.

John, V., Yoneda, K., Liu, Z., Mita, S., 2015. Saliency map generation by the convolutional neural network for real-time traffic light detection using template matching. IEEE Trans. Comput. Imag. 1 (3), 159–173.

Kim, H.-K., Shin, Y.-N., Kuk, S.-g., Park, J.H., Jung, H.-Y., 2013. Night-time traffic light detection based on SVM with geometric moment features. World Academy of Science, Engineering and Technology (WASET), vol. 7, pp. 454–457.

Levinson, J., Askeland, J., Dolson, J., Thrun, S., 2011. Traffic light mapping, localization, and state detection for autonomous vehicles. In: International Conference on Robotics and Automation (ICRA). IEEE, Shanghai, China, pp. 5784–5791. 9–13 May.

Omachi, M., Omachi, S., 2009. Traffic light detection with color and edge information. In: 2nd International Conference on Computer Science and Information Technology (ICCSIT). IEEE, Beijing, China, pp. 284–287. 8–11 August.

Omachi, M., Omachi, S., 2010. Detection of traffic light using structural information. In: 10th International Conference on Signal Processing (ICSP). IEEE, Beijing, China, pp. 809–812. 24–28 October.

Philipsen, M.P., Jensen, M.B., Trivedi, M.M., M?gelmose, A., Moeslund, T.B., 2015. Ongoing work on traffic lights: detection and evaluation. In: 2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 1–6.

Shadeed, W.G., Abu-Al-Nadi, D.I., Mismar, M.J., 2003. Road traffic sign detection in color images. 10th International Conference on Electronics, Circuits and Systems (ICECS), vol. 2. IEEE, Sharjah, United Arab Emirates, pp. 890–893. 14–17 December.

Siogkas, G., Skodras, E., Dermatas, E., 2012. Traffic lights detection in adverse conditions using color, symmetry and spatiotemporal information. In: International Conference on Computer Vision Theory and Applications (VISAPP). Rome, Italy, pp. 620–627. 24–26 February.

Sooksatra, S., Kondo, T., 2014. Red traffic light detection using fast radial symmetry transform. In: 11th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). Nakhon Ratchasima, Thailand, pp. 1–6. 14–17 May.

Tae-Hyun, H., In-Hak, J., Seong-Ik, C., 2006. Advances in Image and Video Technology. Detection of Traffic Lights for Vision-Based Car Navigation System. Springer, Berlin, Heidelberg, pp. 682–691.

Trehard, G., Pollard, E., Bradai, B., Nashashibi, F., 2014. Tracking both pose and status of a traffic light via an interacting multiple model filter. In: 17th International Conference on Information Fusion. IEEE, Salamanca, Spain, pp. 1–7. 7–10 July.

Wang, C., Jin, T., Yang, M., Wang, B., 2011. Robust and real-time traffic lights recognition in complex urban environments. Int. J. Comput. Intell. Syst. 4 (6), 1383–1390.

Yilmaz, A., Javed, O., Shah, M., 2006. Object tracking: a survey. ACM Comput. Surv. 38 (4), 1–45.