TRAFFIC LIGHT DETECTION USING CONIC SECTION GEOMETRY

S. Hosseinyalmdary [a*], Alper Yilmaz[a]

[a] Photogrammetric Computer Vision Laboratory (PCVLab),
Civil and Environmental Engineering and Geodetic Science Department, The Ohio State University
2070 Neil avenue, Columbus, OH, USA 43210 - (hosseinyalamdary.1, yilmaz.15)@osu.edu

Commission I, WG I/Vb

KEY WORDS: Traffic Light Detection, Conic Section, Homography Transformation, OpenStreetMap

ABSTRACT:

Traffic lights detection and their state recognition is a crucial task that autonomous vehicles must reliably fulfill. Despite scientific endeavors, it still is an open problem due to the variations of traffic lights and their perception in image form. Unlike previous studies, this paper investigates the use of inaccurate and publicly available GIS databases such as OpenStreetMap. In addition, we are the first to exploit conic section geometry to improve the shape cue of the traffic lights in images. Conic section also enables us to estimate the pose of the traffic lights with respect to the camera. Our approach can detect multiple traffic lights in the scene, it also is able to detect the traffic lights in the absence of prior knowledge, and detect the traffics lights as far as 70 meters. The proposed approach has been evaluated for different scenarios and the results show that the use of stereo cameras significantly improves the accuracy of the traffic lights detection and pose estimation.

1. INTRODUCTION

The traffic light detection and its state recognition is essential for quickly emerging fully automated vehicles. The state of traffic light includes red, yellow, and green signals that authorizes the right of way and consequently, correctly recognizing the state of traffic light is crucial for safe driving. Although many sensors are applied in fully automated vehicles, the recognition of the state of traffic light is feasible using color cameras on the platform. Thus, we focus on the traffic light detection and its state recognition using color cameras mounted on the platform.

This is still a challenging problem due to the fact that there are many variations of traffic lights. For instance, the shape of traffic lights are not always identical and traffic light lenses can be installed horizontally or vertically. Moreover, size of traffic lights may change from one country to another or from one state to another. Traffic light may be installed on a pole or suspended and they can be located on the right or left side of the road. Moreover, there may be one traffic light for all the lanes or multiple traffic lights for each lane.

In addition, camera sensors are imperfect and the environment may not be benign to capture good images. The images are always noisy especially when low cost cameras are used. Also, motion blurriness deteriorates the image quality when the platform moves. Additionally, camera lens distortion may change geometry of the traffic light if it is not calibrated. In addition, the traffic light has low resolution when it is far from the camera and the low resolution traffic lights may not maintain the shape cue. Furthermore, the lighting situation may create problems that impede correct detection of the traffic lights. For instance, the images can be overexposed during sunrise and sunset when the sun is low. Also, colors are not observed identical in daylight and nightlight.

Knowing these problems, traffic light detection is not a trivial task. In this paper, the color and shape cues are used to detect the traffic lights. In addition, prior knowledge of existence of a traffic light is applied using GIS maps. However, the GIS maps may not be comprehensive and accurate and the traffic lights may not to be accurately localized in the maps. This paper provides an approach to detect traffic lights considering the changes in color and shape cues due to different error sources and inaccurate GIS maps.

In this paper, we apply conic section geometry to detect the traffic lights since conic sections are primitive geometry shapes with multiple properties. Traffic light lens is a circular object which is a conic section. Additionally, conic sections are preserved under perspective geometry, that is, the projection of conic section is a conic section. Conic sections are represented by matrices and conic section geometry is easy to manipulate.

1.1 Previous work

The traffic light detection remains an unsolved problem in fully automated vehicles. More successful works apply prior knowledge such as GIS maps to initialize their search space and improve the traffic light detection results. Levinson et al. (Levinson et al., 2011b, Levinson et al., 2011a) have assumed that the traffic lights can be retrieved from databases such as Google Maps and the search space of traffic lights will be reduced knowing their position. Fairfield and Urmson (Fairfield and Urmson, 2011) have also considered the use of prior maps to predict the location of traffic lights in images. Their GIS maps are fairly accurate and their results show that the use of prior information improves the traffic light detection accuracy. However, these databases are not publicly available in vector format. One of the spatial databases is OpenStreetMap (OSM) which is a vector based GIS database, has world coverage, and is publicly available. This database does not guarantee accurate features and these features may be positioned far from their actual location and even some of the features may be erroneously excluded in the database. OSM has been used in this paper to predict the existence of the traffic lights.

In addition to GIS maps, shape cue can be applied to detect traffic lights. Since the traffic light lens is circular, geometry is a

[*] Corresponding author

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
doi:10.5194/isprsannals-III-1-191-2016

strong constraint to detect the traffic lights. The circular Hough transform is applied to detect the traffic signs (Caraffi et al., 2008, Huang and Lee, 2010) that may be applicable to traffic lights too. However, the use of Hough transform is computationally expensive and it may not be applicable for real time applications. Conic section geometry has been extensively studied for many years and it has been used for stereo, motion, and pose estimation (De Ma, 1993). It has been shown that the pose of the camera can be estimated with respect to the object by observing conic sections of the object (Kannala et al., 2006). In this paper, the use of conic sections obtains a strong geometric constraint to detect the traffic lights and their pose.

Furthermore, color cue plays an important role in every traffic light detection. As previously mentioned, the color of traffic lights can be changed due to different lighting situations. Different color spaces have also been investigated to more robustly detect the traffic lights (John et al., 2014) and multiple exposure images are tested to rigorously detect traffic lights in dark and bright environments (Jang et al., 2014). Diaz-Cabrera et al. (Diaz-Cabrera et al., 2015) claim that color segmentation using fuzzy clustering can improve the traffic light detection results. In this paper, Hue, Saturation, Luminance (HSL) color space is used to detect the traffic lights since it is more resilient to illumination as opposed to Red, Green, Blue (RGB) color space.

The spatial constraint also restrict the position of traffic lights and it can be used to remove false positive detection results that do not follow certain height characteristics. For instance, Siogkas et al. (Siogkas et al., 2012) apply the fact that the traffic light should be above the horizon and they reduce the search space for the traffic lights detection. Finally, characteristics of the traffic light box have been used and several classifiers have been trained to detect these boxes. Wang et al. (Wang et al., 2011) have used the template matching algorithm to detect the traffic light boxes. Also, a few papers, such as (Levinson et al., 2011b, Levinson et al., 2011a), have used machine learning approach to detect the traffic light boxes.

2. METHODOLOGY

This section explains how the visible traffic lights are selected and retrieved from GIS maps. The projection of visible traffic lights into image is not sufficient to initialize search space. Prior knowledge based approaches are not accurate enough when inaccurate GIS maps are used. We apply conic section geometry to detect traffic lights using mono and stereo cameras. The pose of the detected traffic light can be estimated with respect to the camera coordinate system using conic section geometry.

2.1 Visible traffic lights from database

Let's assume that the sensors installed on the platform have been calibrated and the lever arms and boresights of the sensors have been specified. Therefore, the observations are converted to a unified coordinate system which is camera coordinate system here. The integrated GPS/IMU navigation system provides position and orientation of the platform and its solution is accurate as long as the GPS signals are not blocked. Knowing absolute position of the platform enables us to retrieve the information of the nearby traffic lights from available databases such as OpenStreetMap (OSM). Figure 1 shows the position of the platform in blue, the road links in black and the traffic lights in red, for 500 meters within the vicinity of the platform. The neighborhood of the platform is zoomed and the visibility of the camera mounted on the platform is shown by a blue triangle.

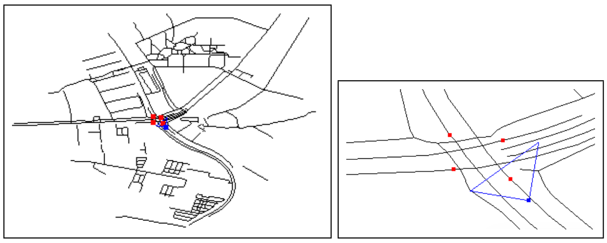

Figure 1: Left: the road links and traffic lights (red points) in the vicinity, 500 meters × 500 meters, of the platform (blue point). Right: the traffic light which is within the region of interest is selected from OpenStreetMap. The region of interest, which is the field of view of the camera up to 70 meters, is shown in blue triangle.

The traffic lights that are visible in the image and are close to the platform should be retrieved from OSM. The absolute position estimated by GPS is used to find the adjacent traffic lights. In addition, the orientation of the platform estimated by IMU and the field of view of the camera, here 90°, are used to find the visible traffic lights.

2.2 Projection of the visible traffic lights into image

If two dimensional position of a traffic light is retrieved from GIS maps, three dimensional position of the traffic light, $\mathbf{X}$, is provided assuming the height of the traffic light is known. In order to estimate the image position of traffic lights, projection matrix, $\mathtt{P}$, should be estimated. The projection matrix maps three dimensional position of the traffic light into two dimensional image space. Two dimensional image coordinates, $\mathbf{x}$, is estimated such that

$$\mathbf{x} = \mathtt{P}\mathbf{X}. \tag{1}$$

The projection matrix can be calculated using rotation matrix and translation vector of the traffic light in the camera coordinate system. Let's assume that $\mathtt{R}_o^c$ and $\mathbf{t}_c$ are the rotation matrix and translation vector between the traffic light coordinate system and the camera coordinate system in the camera coordinate system.

$$\mathtt{P} = \mathtt{K}[\mathtt{R}_c|\mathbf{t}_c]. \tag{2}$$

If the camera is calibrated before the operation, the calibration matrix, $\mathtt{K}$, is known. The rotation matrix can be brought to the camera coordinate system, such that

$$\mathtt{R}_c = \mathtt{R}_i^c \mathtt{R}_n^i \mathtt{R}_n, \tag{3}$$

where $\mathtt{R}_i^c$ is called boresight between the camera and IMU sensors and it is estimated from the calibration procedure. The rotation matrix between the traffic light coordinate system and the camera coordinate system in global coordinate system is $\mathtt{R}_n$. The translation vector, $\mathbf{t}_n$, is the difference of the traffic light position and camera position in the global coordinate system and it can be converted to the translation vector in the camera coordinate system. The translation vector in global coordinate system is transferred to the camera coordinate system such that

$$\mathbf{t}_c = \mathtt{R}_i^c(\mathtt{R}_n^i \mathbf{t}_n + \mathbf{t}_i^c), \tag{4}$$

where $\mathbf{t}_i^c$ is the lever arm between IMU and camera sensors and it is estimated from calibration procedure. If the GIS database and navigation system are fairly accurate, the projected position of the traffic light can be used to initialize the search space of the traffic lights. Generally speaking, the direction and height of the traffic lights are not stored in GIS databases. In addition, the traffic light is projected as a point and its size is not known. Moreover, the information of traffic lights may not be correctly collected and stored in GIS database. Therefore, reducing the search space of the traffic lights may lead to incorrect traffic light detection especially for inaccurate GIS databases such as OSM.

The position of visible traffic lights are retrieved from OSM and projected into image space in Figure 2. There are a few problems if the search space is initialized using the projected traffic light. First, there are multiple traffic lights in the image, whereas the traffic lights are marked as one feature in OSM. Second, the size of the traffic light is not known in the image space and therefore, different scales of the traffic lights should be searched.

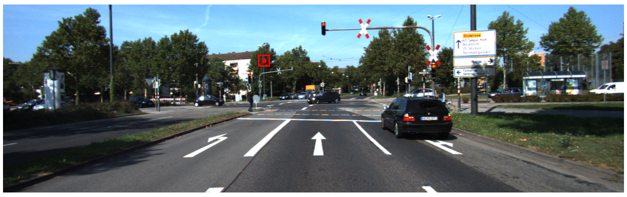

Figure 2: The projection transformation of the traffic light is approximated using the position of the traffic light in OpenStreetMap. The projected traffic light lens into the image space is shown with a circle and a rectangle is drawn around it for better visualization.

2.3 Traffic light detection using mono camera

The traffic light lens is circular on a plane. Therefore, the traffic light lens is a conic section and it follows the conic section geometry. Conic sections are closed form, compact, easy to manipulate shapes that have a symmetric matrix representation. Under perspective geometry, the circular lens will be seen as an ellipse. Let's assume that the circular lens is a conic section, $\mathtt{C}$ and its image is an ellipse which is still a conic section, $\mathtt{C}'$ in the image space. Since the traffic light is assumed to be planar, transformation between the traffic light lens and its image is homography transformation. Therefore, the homography transformation, $\mathtt{H}$, is used to map $\mathtt{C}$ into $\mathtt{C}'$ such that

$$\mathtt{C}' = \mathtt{H}^{-\top} k\mathtt{C}\mathtt{H}^{-1}, \tag{5}$$

Equation (5) estimates the homography transformation up to scale and k is the scale of the homography transformation. Let's assume that the traffic light lens radius is r, the conic section in the traffic light lens plane is a circle represented by a symmetric matrix, such that

$$\mathtt{C} = \begin{bmatrix} \frac{1}{r^2} & 0 & 0 \\ 0 & \frac{1}{r^2} & 0 \\ 0 & 0 & -1 \end{bmatrix}. \tag{6}$$

The rotation matrix, $\mathtt{R}$, is composed of three column vectors, $\mathtt{R} = [\mathbf{r_1}, \mathbf{r_2}, \mathbf{r_3}]$. The homography transformation is calculated such that

$$\mathtt{H} = \mathtt{K}[\mathbf{r_1}\mathbf{r_2}\mathbf{t}_c] = \mathtt{KRT}, \tag{7}$$

where $\mathtt{T} = \begin{bmatrix} 1 & 0 & -t_1 \\ 0 & 1 & -t_2 \\ 0 & 0 & -t_3 \end{bmatrix}$, $\mathbf{t}_o = [t_1 t_2 t_3]^\top$ is the translation vector in the traffic light (object) coordinate system and $\mathbf{t}_c = -\mathtt{R}\mathbf{t}_o$. Replacing the conic section, $\mathtt{C}$, in (6) and homography transformation, $\mathtt{H}$, in (7) into (5) provides conic section in the image space, $\mathtt{C}'$. By inverting (5), the conic section in image space is converted into the conic section in the traffic light coordinate system, such that

$$k\mathtt{C} = \mathtt{H}^\top \mathtt{C}'\mathtt{H}. \tag{8}$$

Replacing (7) into (8), the homography transformation is estimated, such that

$$k\mathtt{C} = \mathtt{T}^\top \mathtt{R}^\top \mathtt{K}^\top \mathtt{C}'\mathtt{KRT}. \tag{9}$$

Let's move $\mathtt{T}$ to the left side of (9) such that

$$k\mathtt{T}^{-\top}\mathtt{C}\mathtt{T}^{-1} = \mathtt{R}^\top \mathtt{K}^\top \mathtt{C}'\mathtt{KR}. \tag{10}$$

If two sides of (10) are equal, their eigenvalues are equal too. If λ_1,λ_2, and λ_3 are the eigenvalues of $\mathtt{K}^\top \mathtt{C}'\mathtt{K}$, it has been shown that (De Ma, 1993)

$$\lambda_1\lambda_2\lambda_3 = \frac{-k^3}{t_3^2 r^4}, \tag{11}$$

$$\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3 = k^2\frac{d^2 - 2r^2}{t_3^2 r^4}, \tag{12}$$

$$\lambda_1 + \lambda_2 + \lambda_3 = k\frac{d^2 + t_3^2 - r^2}{t_3^2 r^2}, \tag{13}$$

where r is the radius of the traffic light lens and $d^2 = t_1^2 + t_2^2 + t_3^2$ is the distance between the camera and traffic light coordinate systems. When t_3 and d are calculated, the choice of t_1 and t_2 is not important since the traffic light lens is a circle and it should only be chosen to the extent that $d^2 = t_1^2 + t_2^2 + t_3^2$.

Two sides of (10) are decomposed to their eigenmatrices and eignevlaues using singular value decomposition, $\mathtt{K}^\top \mathtt{C}'\mathtt{K} = \mathtt{UDU}^\top$ and $\mathtt{T}^{-\top}\mathtt{C}\mathtt{T}^{-1} = \mathtt{VDV}^\top$, where $\mathtt{U}$ and $\mathtt{V}$ are eigenmatrices and $\mathtt{D}$ is a diagonal matrix containing the eigenvalues, λ_1,λ_2, and λ_3. Equation (10) is applied to estimate the rotation matrix, $\mathtt{R}$, such that

$$\mathtt{UDU}^\top = \mathtt{R}^\top \mathtt{VDV}^\top \mathtt{R}. \tag{14}$$

and rotation matrix is calculated in the way that

$$\mathtt{R} = \mathtt{UWV},^{-1} \tag{15}$$

where $\mathtt{W} = \pm\mathtt{I}_{3\times3}$. The rotation matrix and translation vector are calculated from these equations are used to estimate the homography transformation and the estimated homography transformation is used to back-project the conic section in image space into the traffic light coordinate system. Unfortunately, a single conic does not provide enough information to verify if a conic section is a traffic light. In the next section, stereo cameras are used to improve the geometry and provide sufficient constraints to detect the traffic lights.

2.4 Traffic light detection using stereo cameras

If there are two cameras installed on the platform, the traffic lights can be detected using stereo cameras. Let's assume that $\mathtt{R}_L$ and $\mathbf{t}_L$ are rotation matrix and translation vector from the left camera coordinate system to the traffic light coordinate system and $\mathtt{R}_R$ and $\mathbf{t}_R$ are rotation matrix and translation vector from the right camera coordinate system to the traffic light coordinate system. In addition, $\mathtt{R}$ and $\mathbf{t}$ are the rotation matrix and translation vector from the right camera coordinate system to the left camera coordinate system. These parameters, which are called extrinsic camera calibration parameters, can be estimated from the calibration procedure. The rotation matrix and translation vector of left or right camera can be estimated using the rotation matrix and translation vector of the other camera, such that

$$\mathtt{R}_R = \mathtt{R}\mathtt{R}_L, \tag{16}$$

$$\mathbf{t}_R = \mathtt{R}\mathbf{t}_L + \mathbf{t}. \tag{17}$$

similar to (8), the traffic light lens and its projection into left and right images are written, such that

$$k_L\mathtt{C} = \mathtt{H}_L^\top\mathtt{C}'_L\mathtt{H}_L, \tag{18}$$

$$k_R\mathtt{C} = \mathtt{H}_R^\top\mathtt{C}'_R\mathtt{H}_R. \tag{19}$$

If the 2×2 upper left submatrix of (18) and (19) are chosen, these two equations will be simplified, such that

$$k_L(\mathtt{C})^{2\times2} = (\mathtt{R}_L^\top\mathtt{K}_L^\top\mathtt{C}'_L\mathtt{K}_L\mathtt{R}_L)^{2\times2}, \tag{20}$$

$$k_R(\mathtt{C})^{2\times2} = (\mathtt{R}_R^\top\mathtt{K}_R^\top\mathtt{C}'_R\mathtt{K}_R\mathtt{R}_R)^{2\times2}. \tag{21}$$

Since the left side of (20) and (21) are equal up to a scale, these equations can be reduced such that

$$[\mathtt{R}_L^\top(\mathtt{K}_L^\top\mathtt{C}'_L\mathtt{K}_L - \frac{k_L}{k_R}\mathtt{R}^\top\mathtt{K}_R^\top\mathtt{C}'_R\mathtt{K}_R\mathtt{R})\mathtt{R}_L]^{2\times2} = \mathtt{0}^{2\times2} \tag{22}$$

It can be proven that if the 2×2 upper left matrix of a 3×3 matrix is zero, the determinant of the 3×3 matrix is zero. Let's assume the inner part of (22) is $\mathtt{B}$ matrix, where $\mathtt{B} = \mathtt{K}_L^\top\mathtt{C}'_L\mathtt{K}_L - \frac{k_L}{k_R}\mathtt{R}^\top\mathtt{K}_R^\top\mathtt{C}'_R\mathtt{K}_R\mathtt{R}$. The determinant of this matrix is zero, $\det(B) = 0$.

Therefore, one of the eigenvalues of matrix $\mathtt{B}$ is zero. Let's assume λ_1 and λ_2 are non-zero eigenvalues of matrix $\mathtt{B}$ and its corresponding eigenvectors are $\mathbf{s}_1$ and $\mathbf{s}_2$. The third column vector of the rotation matrix is estimated such that

$$\mathbf{r}_3 = \pm\text{norm}(\sqrt{\lambda_1}\mathbf{s}_1 \pm \sqrt{\lambda_2}\mathbf{s}_2). \tag{23}$$

Let's construct matrix $\mathtt{A} = \mathtt{K}_L^\top\mathtt{C}'_L\mathtt{K}_L - \mathbf{r}_3\mathbf{r}_3^\top\mathtt{K}_L^\top\mathtt{C}'_L\mathtt{K}_L$. One of the eigenvalues of matrix $\mathtt{A}$ is zero since $\det\mathtt{A} = 0$. The two eigenvectors corresponding to non-zero eigenvalues are the first and second column vectors of rotation matrix, $\mathbf{r}_1$ and $\mathbf{r}_2$. When the rotation matrix, $\mathtt{R}_L = [\mathbf{r}_{L1}\,\mathbf{r}_{L2}\,\mathbf{r}_{L3}]^\top$, the rotation matrix $\mathtt{R}_R$ is constructed using (16). The translation vector is the solution the equations system

$$\begin{cases} \mathbf{t}_L^\top\mathtt{C}'_L\mathbf{r}_{L1} = 0 \\ \mathbf{t}_L^\top\mathtt{C}'_L\mathbf{r}_{L2} = 0 \\ \mathbf{t}_R^\top\mathtt{C}'_R\mathbf{r}_{R2} = 0 \\ \mathbf{t}_R = \mathtt{R}\mathbf{t}_L + \mathbf{t}, \end{cases} \tag{24}$$

where $\mathbf{r}_{R2}$ is the second row vector of $\mathtt{R}_R$.

2.5 Conic matching between epochs

If the traffic light is visible in two consecutive epochs, A corresponding conic section should be matched in two epochs. Let's assume that $\mathtt{C}'_t$ and $\mathtt{C}'_{t+1}$ are two conic sections at time t and $t+1$ and it should be verified whether these conics belong to one object and therefore, these are correspondence. If the image points $\mathbf{x}_t$ and $\mathbf{x}_{t+1}$ lay on the conic sections, such that

$$s_t\mathbf{x}_t^\top\mathtt{C}'_t\mathbf{x}_t = 0, \tag{25}$$

$$s_{t+1}\mathbf{x}_{t+1}^\top\mathtt{C}'_{t+1}\mathbf{x}_{t+1} = 0, \tag{26}$$

If these conic sections are correspondence, there is a homography transformation between the image points of these conics, such that

$$s\mathbf{x}_t = \mathtt{H}_{t+1}^t\mathbf{x}_{t+1}, \tag{27}$$

where s is a scale factor. It can be shown that the homography transformation, $\mathtt{H}_{t+1}^t$, is estimated based on the rotation matrix and translation vector, such that

$$\mathtt{H}_{t+1}^t = k(\mathtt{R}_{t+1}^t + \frac{\mathbf{t}_{t+1}^t\mathbf{n}^\top}{d}), \tag{28}$$

where $\mathtt{R}_{t+1}^t$ is the rotation matrix between two images and it can be estimated from IMU sensor and converted to image coordinate system. Similarly, translation vector, $\mathbf{t}_{t+1}^t$, can be estimated GPS sensors and this vector is transferred into the image coordinate system. If the normal vector to the traffic light plane, $\mathbf{n}$, and the distance of traffic light to the image, d, are known from previous epoch, the homography transformation is estimated up to the scale k.

If () is replaced into (), the conic section at time t, $\mathtt{C}'_t$, is related to conic section at time $t+1$, $\mathtt{C}'_{t+1}$, such that

$$\mathbf{x}_{t+1}^\top\mathtt{H}_{t+1}^\top\mathtt{C}'_{t+1}\mathtt{H}_{t+1}\mathbf{x}_{t+1} = c\mathtt{C}'_t, \tag{29}$$

Two sides of these equations are known up to scale, c. Therefore, the scale factor can be estimated, such that

$$c = \frac{\det\left(\mathbf{x}_{t+1}^{\top}\mathtt{H}_{t+1}^{\top}\mathtt{C}'_{t+1}\mathtt{H}_{t+1}\mathbf{x}_{t+1}\right)}{\det\left(\mathtt{C}'_{t}\right)}, \tag{30}$$

when scale factor, c, is determined, the conic section at time t is converted into time $t+1$ and if the difference of the converted conic and observed conic at time $t+1$ is less than a threshold, ϵ, these conics are matched between time t and $t+1$, such that

$$||\mathbf{x}_{t+1}^{\top}\mathtt{H}_{t+1}^{\top}\mathtt{C}'_{t+1}\mathtt{H}_{t+1}\mathbf{x}_{t+1} - c\mathtt{C}'_{t}|| < \epsilon \tag{31}$$

2.6 Temporal constraint

Let's divide a conic section state into two classes of traffic lights, o_1, and others, o_2. Let's assume that $p(C_t = o_1)$ is the probability of the conic is a traffic light at time t and $p(C_t = o_2) = 1 - p(C_t = o_1)$ otherwise. If the shape and color of the traffic light is observed at time t, the probability of the traffic light given the observed shape and color cue, Z_t, is $p(C_t = o_1|Z_t)$. In addition, if a conic section is labeled as traffic light in the previous epochs, it is most likely to be labeled as traffic light at this epoch. Therefore, the probability of the traffic light detection at time t is $p(C_t = o_1|Z_t, C_{1:t-1} = o_1)$. By Markov a

The traffic light detection can be modeled as Hidden Markov Model (HMM), such that

$$p(C_t|Z_t, C_{1:t-1}) = p(Z_t|C_{1:t})\frac{p(C_t|C_{1:t-1})}{p(Z_t|C_{1:t-1})} \tag{32}$$

Since observation at time t, Z_t, is independent from the state of the conic at time $1 : t-1$, therefore $p(Z_t|C_t = o_1) = p(Z_t|C_{1:t} = o_1)$ and $p(Z_t) = p(Z_t|C_{1:t-1} = o_1)$. In addition, $p(Z_t)$ is a normalization factor, α and does not have impact on the probability. Therefore, this equation can be shortened, such that

$$p(C_t|Z_t, C_{1:t-1}) = \alpha p(Z_t|C_t)p(C_t|C_{1:t-1}) \tag{33}$$

By the Markov assumption, the state of the conic only depends on the previous state of the conic, such that

$$p(C_t|C_{1:t-1}) = p(C_t|C_{t-1}) \tag{34}$$

If the traffic light is detected in the previous epoch and the pose of the camera is accurately estimated, the location of the traffic light can be transferred into the next epoch. If the platform does not move, the location of the traffic light in the previous epoch does not change in the next epoch and $p(C_t = o_1|C_{t-1} = o_1) = 1$. However, the pose of the camera between two epoch is estimated using GPS/IMU navigation solution. Therefore, the transition term, $p(C_t|C_{t-1})$, depends on the accuracy of the navigation solution. Let's assume the position and orientation errors δp and $\delta\theta$, the transition condition is estimated such that

$$p(C_t|C_{t-1}) = \frac{1}{(2\pi)^2 s_{\delta p}^2 s_{\delta\theta}^2}\exp(-\frac{1}{2}(\frac{\delta p}{s_{\delta p}^2} + \frac{\delta\theta}{s_{\delta\theta}^2})) \tag{35}$$

2.7 Evidence

The conic section geometry provides strong constraint on the shape of the traffic light lenses. However, there are other observations that may be applied to detect the traffic lights. The color of the traffic light lens is limited to red, yellow, and green. Unfortunately, the color may be changed due to different illuminations and distance of the traffic light. Also, the box of the traffic light is applied since it has a standard shape. If the traffic light lens is detected correctly, the box of the traffic light can be reconstructed. The observed box and reconstructed box should be in agreement. Otherwise, the object may be a red, yellow or green light located on the different objects such as cars.

Therefore, the observation is actually a set of color, shape, box cues and depth, $Z_t = \{z_c, z_s, z_b, d\}$. Unfortunately, all of these observations are correlated. The box reconstruction highly depends on the shape of the traffic light lens. The shape of the traffic light lens depends on the color cue and the color cue depends on the depth of the traffic light. Therefore, the probability of these observations is

$$\begin{gathered} p(Z_t|C_t) = p(z_c \cap z_s \cap z_b \cap d|C_t) = \\ p(z_b|z_c, z_s, d, C_t)p(z_s|z_c, d, C_t)p(z_c|d, C_t)p(d|C_t) \end{gathered} \tag{36}$$

The depth of the object does not depth on its label and therefore, $p(d|C_t) = p(d)$. The probability of the measure depth merely depends on the accuracy of the depth measurement. Therefore, if the stereo images are used for the depth estimation, the probability of the measured depth depends on the accuracy of stereo matching. The laser scanner is used, the accuracy of the laser scanner indicates the accuracy of the measured depth.

In order to calculate the probability of the color of the traffic light lens, the image color space is converted from Red,Green, and Blue (RGB) into Hue, Saturation, and Luminance (HSL). The HSL color space is more resilient against different illuminations. The less saturation and luminance the darker and less likely the traffic light lens. Here, we use logistic function for the probablity distribution function of the saturation of luminance.

$$\begin{cases} p(S) = \frac{1}{1+\exp(-k_s(x-x_s))} \\ p(V) = \frac{1}{1+\exp(-k_v(x-x_v))} \end{cases} \tag{37}$$

where k_s and k_v are the steepness of the logistic function curves, x_s and x_v are the midpoint of the logistic functions and these parameters are related to the measured depth of the traffic light.

For the hue component, we assume that the probability of a hue component is the summation of the probability of the red, yellow, and green colors. Obviously, the probability distribution function should be normalized. Adding the hue, saturation, and luminance of the color space, the probability of the color cue for a traffic light is estimated such that

$$p(H) = \frac{1}{3}[p(H = r) + p(H = y) + p(H = g)] \tag{38}$$

where $p(H = r)$, $p(H = y)$, and $p(H = g)$ are the probabilities of the traffic light in red, yellow, and green and these probabilities are estimated based on the histogram of the traffic light lens color.

$$p(Z_c|Ct = o_1) = p(H)p(S)p(V) \tag{39}$$

The traffic light lens is a circle with the known radius, r. Therefore, the observed ellipse in the image space can be back-projected into the traffic light space and the back-projected ellipse may be still an ellipse due to different error sources. Let's assume $\hat{a}$ and $\hat{b}$ are the semi-major and semi-minor axes of the estimated ellipse and they follow normal distribution, such that

$$\begin{aligned}\hat{a} &\sim \mathcal{N}(r, \sigma_a^2)\\ \hat{b} &\sim \mathcal{N}(r, \sigma_b^2),\end{aligned} \tag{40}$$

where σ_a^2 and σ_b^2 are the variances of $\hat{a}$ and $\hat{b}$. Since the variances are not know, it can be estimated and the estimated variances of $\hat{a}$ and $\hat{b}$ are s_a^2 and s_b^2. The covariance between $\hat{a}$ and $\hat{b}$ is ignored since these values are orthogonal. The probability of the observed ellipse is estimated, such that

$$p(z_s|z_c, d, C_t) = \frac{1}{(2\pi)^2 s_a^2 s_b^2} \exp\left(-\frac{1}{2}\left(\frac{\hat{a}-r}{s_a^2} + \frac{\hat{b}-r}{s_b^2}\right)\right) \tag{41}$$

The geometry of the box of the traffic light follows the standards of the transportation and highways administrations. Therefore, if one of the lenses of the traffic light is accurately detected, other lenses is determined since the traffic lights are installed horizontally or vertically. If the other lenses are not led on, they should be appeared black in the image. Therefore, other objects such as tail lights of the cars that do not follow the geometry of the traffic light box, have low probability to be detected as traffic lights.Without loss of generality, let's assume the red lens is led on, the probability of the box is estimated such that

$$\begin{aligned}p(z_b|z_s, z_c, d, C_t) &= p(l2 = \mathit{yellow} \cap l3 = \mathit{green}|l1 = \mathit{red}) =\\ &p(l2 = \mathit{yellow}|l1 = \mathit{red})p(l3 = \mathit{green}|l1 = \mathit{red}).\end{aligned} \tag{42}$$

The probability of the yellow of green lenses are not active is a logistic function, such that

$$\begin{cases} p(l2 = Y|l1 = R) = \frac{1}{1+\exp(-k_Y(x - x_Y))}\\ p(l2 = G|l1 = R) = \frac{1}{1+\exp(-k_G(x - x_G))}\end{cases} \tag{43}$$

where k_Y and k_G are the steepness of the logistic function curves, x_Y and x_G are the midpoint of the logistic functions and these parameters.

3. IMPLEMENTATION

The projected traffic light may not be helpful since position of the traffic light is not accurate in Figure 2. Therefore, the whole image is searched for traffic lights. Since traffic lights have red, yellow, and green circular lenses, the color and geometry cues of these lenses are used to detect traffic lights. Although use of color is not enough to detect the traffic lights, it narrows the search space by removing many objects such as ground which is not red, yellow, or green. In order to find red, yellow, or green color objects, image is converted to Hue, Saturation, Luminance (HSL) color space. The acceptable red, yellow and green thresholds are given in (46). It should be noted that the range of hue is between 0 and 180, and the range of saturation and luminance are between 0 and 255, here. Saturation less than 10 and luminance less than 50 are not acceptable in our algorithm since the objects are too dark with these saturation and luminance.

$$\begin{cases}\text{if } 14 > h \text{ or } 170 < h & \text{then color = red}\\ \text{if } 13 < h < 30 & \text{then color = yellow}\\ \text{if } 30 < h < 100 & \text{then color = green.}\end{cases} \tag{44}$$

The results of the red, yellow, and green objects are shown in Figure 3. However the traffic light lenses are red, yellow, and green, these colors may be shifted in the image. For instance, the green color of the traffic light may be seen as light blue and therefore, the thresholds in (44) are designed in a way that it includes other possible colors of the traffic lights such as light blue. The sky in Figure 3 is chosen as green object for the same reason. In addition, the boundary between red and yellow colors is ambiguous and the red traffic light lenses may be seen as dark yellow lenses and vice versa. Therefore, the red and yellow colors have an overlap in hue.

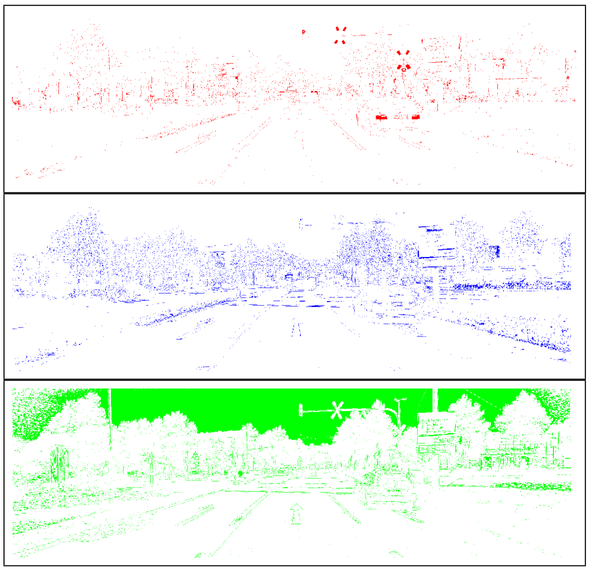

Figure 3: The red, yellow, and green objects are selected using threshold on the hue component of HSL color space. Yellow channel is shown in blue for better visualization.

Noise in color, which is unavoidable in inexpensive cameras, can create salt and pepper noise after the color mask is applied. A median filter is used to remove these fictitious red, yellow, and green pixels. The filtered image when the median filter is applied is shown in Figure 4.

The shape cue of the traffic light lenses can also also be used. The lenses of traffic lights are circular and each circle can be seen an ellipse under perspective geometry. Therefore, objects that their are red, yellow, or green and their geometry is similar to ellipse are chosen as the possible traffic lights. For every red, yellow, and green object the boundaries of the object is extracted and the points on the boundaries are used in a least squares to estimate the best fitted ellipse (Fitzgibbon and Fisher, 1999). If an object is not ellipse shape object, the residuals of the ellipse fitting will be high and therefore, the residuals of the ellipse fitting are used to reject non-ellipse objects. Figure 5 shows the fitted ellipses to red, yellow, and green objects.

Although the color and shape cues have been used, there are many objects in Figure 5 that are not traffic light lenses. For instance, the tail light of the car is a red object which has a shape similar to

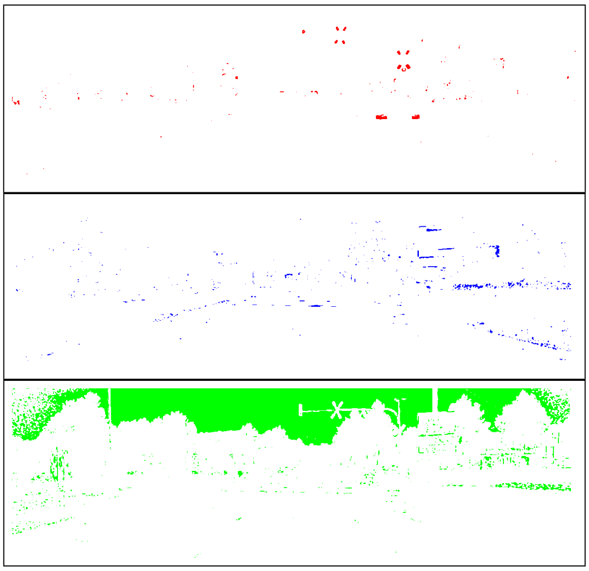

Figure 4: A median filter is applied to the red, yellow and green objects to remove noise from these objects. Yellow channel is shown in blue for better visualization.

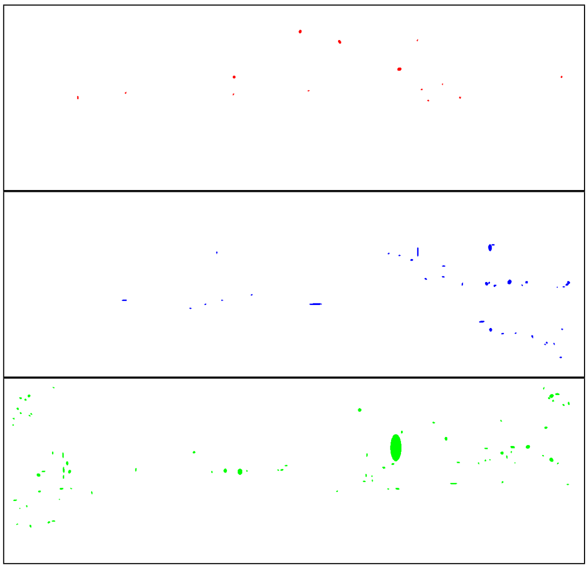

Figure 5: An ellipse is fitted to the boundaries of the connected red, yellow, and green pixels. The objects with large residuals of the ellipse fitting are not ellipse-shape objects and should be removed.

an ellipse. These objects should be detected and removed to detect the traffic lights. The homography transformation in (7) can be used to transfer the red, yellow, and green ellipses in Figure 5 from the image space into the traffic light coordinate system. This back-projection is performed using homography transformation, $\mathtt{H}$. Ideally, $\hat{\mathtt{C}}$ should be equal to the conic section $\mathtt{C}$ in (6). However, the back-projected ellipse will not be a circle due to noise in color image and inaccurate homography transformation. Therefore, $||\hat{\mathtt{C}} - \mathtt{C}||$ can be used to detect the traffic lights.

In addition, the image geometry is enforced that the traffic lights must be in front of the camera and therefore, $t_z > 0$. In addition, the traffic light installation is enforced that the traffic light should be above the ground in a certain height. The traffic light position constraints are used to ensure these limitations in this paper, such that

$$\begin{cases} 3m < t_u < 10m \\ 1m < t_z < 70m, \end{cases} \tag{45}$$

where t_z is the distance of the traffic light in the camera principal axis direction and t_u is the traffic light height above the ground.

Figure 6 shows the traffic light lenses detected in the image using mono camera. The results are not acceptable since there are other objects that are incorrectly labeled as the traffic light lenses. There are two main error sources that leads to incorrect traffic light lens detection: The color is not a robust source of information and it can change in different illumination environments and optical instruments, especially in inexpensive cameras; The shape cue may not be accurate for far traffic lights where these traffic lights have low resolution.

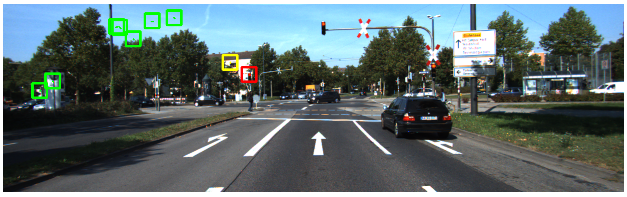

Figure 6: The traffic light lenses are detected using mono camera. There are many false positive traffic light lenses and this approach is not sufficiently robust approach.

3.1 using stereo to improve robustness

In the previous section, it has been shown that the traffic light lenses cannot be robustly detected using one image since there are not sufficient geometrical constraints. Therefore, the stereo cameras are used to provide redundant information and therefore, improve the accuracy of the traffic light detection. If the possible traffic lights are determined using mono camera approach for left and right images independently, the conics in two images should be matched and the pose of each possible traffic light should be estimated using a pair of conics. Let's assume that the rotation matrix and translation vector of a traffic light in left and right images are $\mathtt{R}_L$, $\mathtt{R}_R$, $\mathbf{t}_L$, and $\mathbf{t}_R$. If these rotation matrices and translation vectors satisfy (18) and (19) in a certain level of accuracy, the conic in left and right images are matched. These criteria allows multiple conics allocations and the wrong matches should be removed in further steps. Figure 7 illustrates the matching conics of two images. It should be noted and the red, yellow and green conics should be matched to the similar color conics.

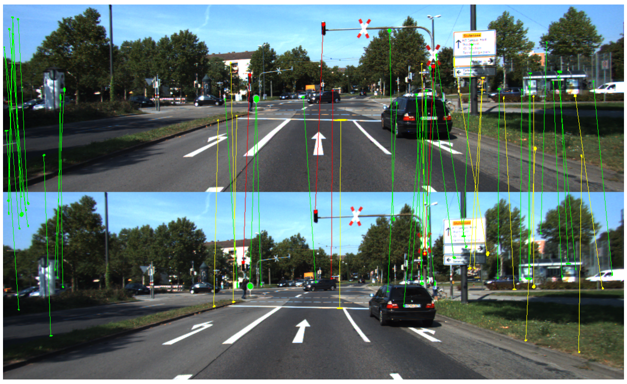

Figure 7: The red, yellow and green ellipses are matched in left and right images.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
doi:10.5194/isprsannals-III-1-191-2016

The approach in section 2.4 can be used to estimate the pose of the conics and validate whether the object is a traffic light. In this paper, those equations are approximated using the intersection approach. If the two ellipses are matched, the center of these ellipses are used to estimate three dimensional position of the traffic light. Let's assume the center of the fitted ellipse of a traffic light lens in the left and right images are $\mathbf{x_L} = [p_x\ p_y]^\top$ and $\mathbf{x_R} = [q_x\ q_y]^\top$.The position of the traffic light in camera coordinate system, $\mathbf{X}$, is estimated using the disparity to distance matrix, $\mathtt{Q}$, such that

$$\hat{\mathbf{X}} = \mathtt{Q}\mathbf{x}' \tag{46}$$

where $\mathbf{x}' = [p_x\ p_y\ dx\ 1]^\top$ and $dx = p_x - q_x$. The parallax in y direction, $dy = p_y - q_y$ should be close to zero if the cameras are aligned in x direction. The disparity to distance matrix, $\mathtt{Q}$, is calculated such that

$$\mathtt{Q} = \begin{bmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 0 & f \\ 0 & 0 & -\frac{1}{b} & -\frac{\delta c_x}{b} \end{bmatrix}, \tag{47}$$

where $\mathbf{c} = [c_x\ c_y]^\top$ and f are the principal point of the image and focal length of the camera and these parameters are calculated from the calibration procedure. The distance between the principal points of two cameras is baseline b and δc_x is the difference of the principal points of the left and right images in x direction. When three dimensional coordinates of the object is estimated, the homography transformation in (7) can be updated such that

$$\hat{\mathtt{H}} = \mathtt{K}[\mathbf{r_1}\mathbf{r_2}\bar{\mathbf{X}}], \tag{48}$$

where $\bar{\mathbf{X}}$ is inhomogeneous coordinates of the object and $\hat{\mathbf{X}} = [\bar{\mathbf{X}}^\top\ 1]^\top$. The homography transformation can be estimated for left image, $\hat{\mathtt{H}}_L$, and right image, $\hat{\mathtt{H}}_R$, if the information of the left camera or right camera are used. If the estimated homography transformation, $\hat{\mathtt{H}}$, is replaced in (8), the estimated traffic light lens in traffic light coordinate system, $\hat{\mathtt{C}}$ will be much more accurate than the one estimated using mono camera.

Figure 8 shows the detected traffic lights using stereo cameras. Two of the three traffic lights are correctly detected and there is no object that incorrectly labeled as the traffic light lens. The state of the traffic lights is correctly recognized and the traffic lights are correctly colored. However, one of the traffic light lenses is not correctly detected. The traffic light lens is missed because its shape is strongly distorted. The resolution of the projected traffic light is low since the traffic light is far from the camera.

4. RESULTS

The images that have been used to describe the methodology and the ones that are used in this section are chosen from the KITTI dataset. Two color 1.4 Megapixel PointGray Flea2 cameras are used to detect the traffic lights. The position of the platform observed by an OXTS RT 3003 integrated GPS/IMU navigation system is used to retrieve the traffic lights from GIS maps (Geiger et al., 2012, Geiger et al., 2013). The calibration procedure has been performed and the calibration parameters such as lens focal length and principal points have been estimated. In addition, the lever arm and boresight of the sensors have been determined with respect to the platform coordinate system (Hosseinyalamdary and Yilmaz, 2014, Hosseinyalamdary et al., 2015).

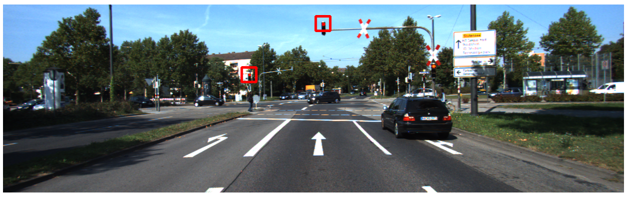

Figure 8: The detected traffic light lenses are correctly detected and their status are correctly recognized using stereo cameras. Although, there is no false positive, there is a missed traffic light lens and it is the result of strong distortion in the shape of traffic light lens.

One of the KITTI scenarios has been used in section 3. to describe the implementation procedure. In addition, three more scenarios are studied in this section and the results are given in Figures 9, 10, and 11. The first row of every figure is the original image taken from the left color camera. The second row is the masked red, yellow and green objects in the left image. The median filter has been applied to remove noise in color cue, similar to Figure 4. The third row is the traffic light detection algorithm using mono camera. The fourth row shows the results of the traffic light detection using stereo cameras.

In Figure 9, the traffic light signal is green. Surprisingly, the green traffic light signal is actually light blue and the hue threshold in (44) should be selected in the way that the light blue is selected as possible green traffic light lens. If the threshold is extended to select the light blue, sky is most likely chosen as the possible traffic light and it should be removed in further processes such as ellipse fitting. In addition, the shade of the objects may be chosen as green object since the boundaries of the shade may be seen as light blue. In mono camera the traffic light lens is correctly detected and its status is correctly recognized. However, many objects are incorrectly labeled as the traffic light lenses. In stereo camera approach, no other object is incorrectly labeled as the traffic light lens while the traffic light lens is correctly selected. It can be concluded that using stereo camera significantly improves the accuracy of the traffic light detection.

Figure 10 shows color of the objects can be changed due to illumination. In this image, sky is captured in white since it is too bright, the trees are captured in black since trees are too dark and there are shades on the street that are captured as green. Due to significant illumination changes in the image, mono camera approach is not able to detect the traffic light correctly and a false positive (shade on the road) is incorrectly labeled as the traffic light. The stereo camera approach detects the traffic light correctly although there is a false positive in this approach, too.

Figure 11 is a challenging scenario since the traffic lights are far from the camera and their projection into image contains few pixels. In addition, these traffic lights are missed in OSM and no prior knowledge of these traffic lights is available. The mono camera based traffic light detection has many false positives that are labeled as traffic light. In stereo camera approach there are still few false positives and the stereo camera algorithm was not able to remove these false traffic lights. These false positives are removed when the camera becomes closer to the traffic lights and resolution of the traffic lights increases in the image.

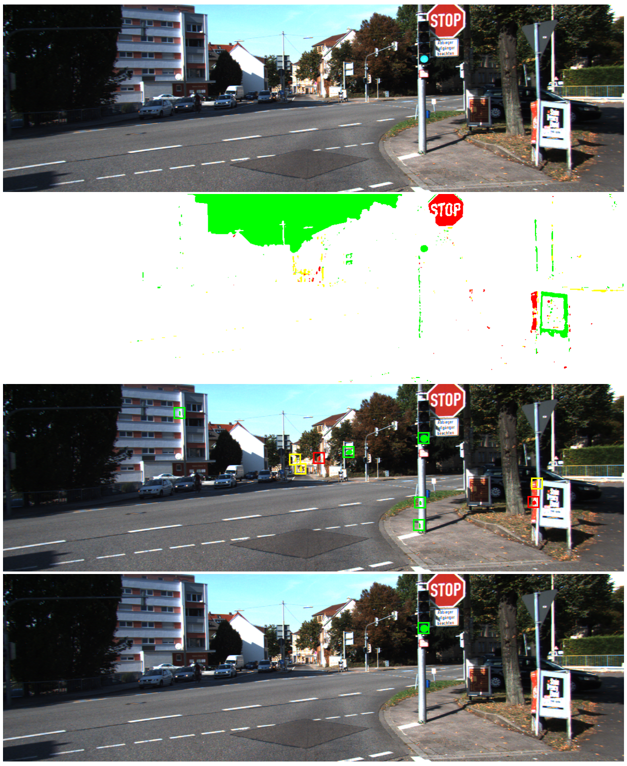

Figure 9: The traffic light detection algorithm using mono camera (third row) has a few false positives. The stereo cameras approach (fourth row) does not contain any false positive while it it maintains the true positives.

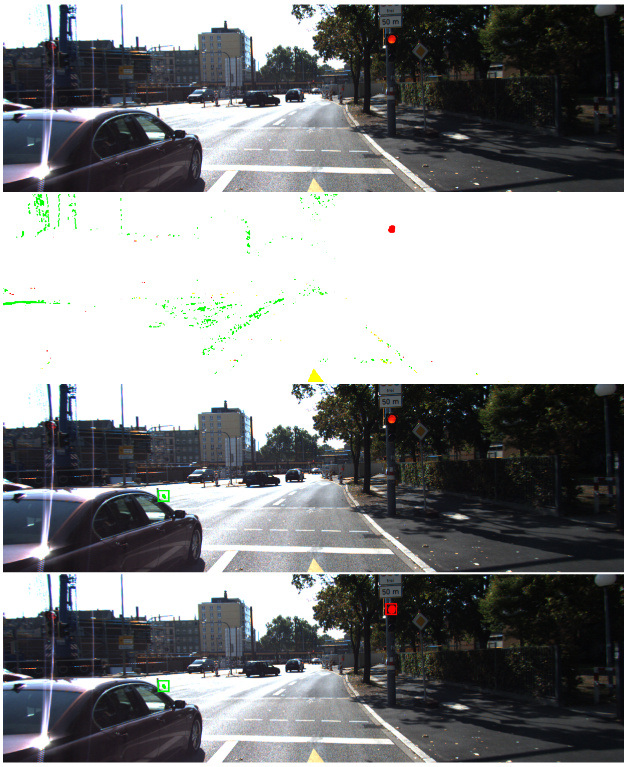

Figure 10: The traffic light detection algorithm using mono camera (third row) can not correctly detect the traffic light and a false positive incorrectly labeled as traffic light. The stereo cameras approach (fourth row) correctly detects the traffic light although a false positive is detected too.

5. CONCLUSION

In this paper, the traffic light detection has been studied and two approaches, mono camera based approach and stereo cameras based approach, are proposed. The color and shape of the traffic light lens and its geometry with respect to the road and the platform are used to distinguish the traffic lights from other objects in the scene. The proposed approaches also investigated the use of prior knowledge in the GIS maps such as OpenStreetMaps to detect the traffic lights. The mono camera based approach is not as robust as the stereo cameras base approach and usually there are a few false positives and false negatives in this approach. It is suggested that a sequence of images are used to improve the robustness of the proposed approach in future.

REFERENCES

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

De Ma, S., 1993. Conics-based stereo, motion estimation, and pose determination. International Journal of Computer Vision 10(1), pp. 7–25.

Diaz-Cabrera, M., Cerri, P. and Medici, P., 2015. Robust realtime traffic light detection and distance estimation using a single camera. Expert Systems with Applications 42(8), pp. 3911–3923.

Fairfield, N. and Urmson, C., 2011. Traffic light mapping and detection. In: The International Conference on Robotics and Automation (ICRA).

Fitzgibbon, A. W.and Pilu, M. and Fisher, R. B., 1999. Direct least-squares fitting of ellipses. 21(5), pp. 476–480.

Geiger, A., Lenz, P. and Urtasun, R., 2012. Are we ready for autonomous driving? the kitti vision benchmark suite. In: Conference on Computer Vision and Pattern Recognition (CVPR).

Geiger, A., Lenz, P., Stiller, C. and Urtasun, R., 2013. Vision meets robotics: The kitti dataset. International Journal of Robotics Research (IJRR).

Hosseinyalamdary, S. and Yilmaz, A., 2014. Motion vector field estimation using brightness constancy motion vector field estimation using brightness constancy assumption and epipolar geometry constraint. In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. II-1.

Hosseinyalamdary, S., Balazadegan, Y. and Toth, C., 2015. Tracking 3d moving objects based on gps/imu navigation solution, laser scanner point cloud and gis data. ISPRS International Journal of Geo-Information 4(3), pp. 1301.

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

Jang, C., Kim, C., Kim, D., Lee, M. and Sunwoo, M., 2014. Multiple exposure images based traffic light recognition. In: Intelligent Vehicles Symposium Proceedings, 2014 IEEE, pp. 1313–1318.

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

Kannala, J., Salo, M. and Heikkilä, J., 2006. Algorithms for computing a planar homography from conics in correspondence. In:

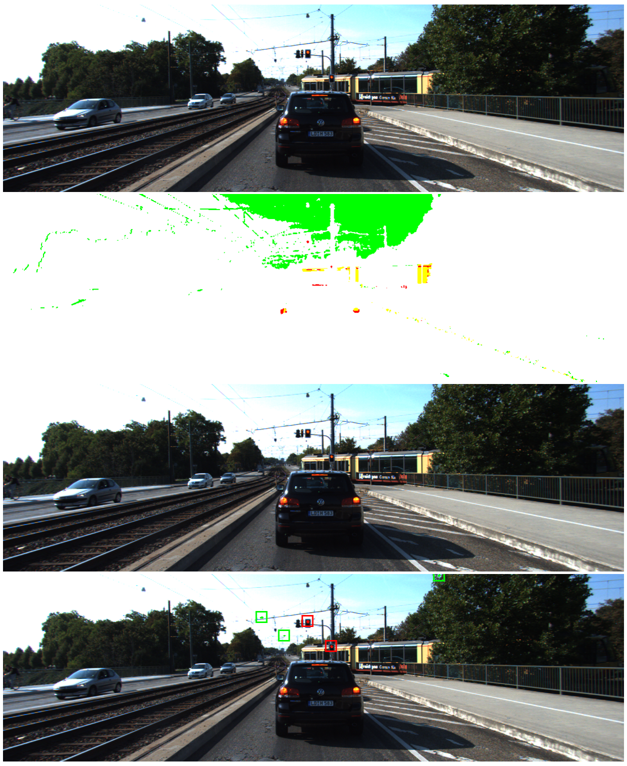

Figure 11: The traffic light detection algorithm is evaluated for a challenging scenario where the traffic light is far from the camera and their resolution is low in the images. Mono camera based approach (third row) is not able to correctly detect the traffic lights are there are a few false positives. The stereo camera (fourth row) can correctly detect the traffic lights however there are few false positives.

M. J. Chantler, R. B. Fisher and E. Trucco (eds), British Machine Vision Conference (BMVC), pp. 77–86.

Levinson, J., Askeland, J., Becker, J., Dolson, J., Held, D., Kammel, S., Kolter, J. Z., Langer, D., Pink, O., Pratt, V., Sokolsky, M., Stanek, G., Stavens, D., Teichman, A., Werling, M. and Thrun, S., 2011a. Towards fully autonomous driving: systems and algorithms. In: Intelligent Vehicles Symposium (IV).

Levinson, J., Askeland, J., Dolson, J. and Thrun, S., 2011b. Traffic light mapping, localization, and state detection for autonomous vehicles. In: International Conference on Robotics and Automation (ICRA).

Siogkas, G., Skodras, E. and 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.

Wang, C., Jin, T., Yang, M. and Wang, B., 2011. Robust and real-time traffic lights recognition in complex urban environments. International journal on computational intelligence techniques, methods and applications 4-6, pp. 1383–1390.