

# Geolocation estimation of target vehicles using image processing and geometric computation



Elnaz Namazi<sup>a,\*</sup>, Rudolf Mester<sup>a</sup>, Chaoru Lu<sup>b</sup>, Jingyue Li<sup>a</sup>

<sup>a</sup> Department of Computer Science, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

<sup>b</sup> Department of Civil Engineering and Energy Technology, Oslo Metropolitan University (OsloMet), Oslo, Norway

## ARTICLE INFO

### Article history:

Received 24 March 2021

Revised 25 October 2021

Accepted 31 October 2021

Available online 8 May 2022

### Keywords:

Target-vehicle localization

Vision-based localization

Geometric computation

GPS coordinate system

Mixed traffic

## ABSTRACT

Estimating vehicles' locations is one of the key components in intelligent traffic management systems (ITMSs) for increasing traffic scene awareness. Traditionally, stationary sensors have been employed in this regard. The development of advanced sensing and communication technologies on modern vehicles (MVs) makes it feasible to use such vehicles as mobile sensors to estimate the traffic data of observed vehicles. This study aims to explore the capabilities of a monocular camera mounted on an MV in order to estimate the geolocation of the observed vehicle in a global positioning system (GPS) coordinate system. We proposed a new methodology by integrating deep learning, image processing, and geometric computation to address the observed-vehicle localization problem. To evaluate our proposed methodology, we developed new algorithms and tested them using real-world traffic data. The results indicated that our proposed methodology and algorithms could effectively estimate the observed vehicle's latitude and longitude dynamically.

© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

A growing body of literature recognizes the importance of intelligent traffic management systems (ITMSs) to manage traffic safely and efficiently. ITMSs mainly rely on traffic data to enhance traffic scene awareness and make smart decisions [1].

There are two main approaches to collecting traffic data for ITMSs. The first approach is based on stationary sensors placed toward road networks, such as inductive loop detectors (e.g., [2]) and closed-circuit television cameras (e.g., [3]). Although this approach is nowadays widely applied to collect traffic data, installing and maintaining these sensors to provide an acceptable coverage range on all roads might be costly [4]. The second approach is based on using modern vehicles (MVs) equipped with sensors. An MV with sensing and communication abilities can collect traffic data mainly about itself and transfer it based on, in general, vehicle-to-vehicle (V2V) or vehicle-to-infrastructure (V2I) communications. To collect enough traffic data with this approach, most vehicles in the traffic need to be MVs with an advanced sensor mounted. However, converting most vehicles into MVs is time-consuming. Studies predict that only 50% of vehicles in the United States will have autonomy in Level 4 (vehicles in Level 4, based on the Society of Automotive Engineers (SAE), have high automation,

with which the automated driving features can drive the vehicle under limited conditions, and the driver holds control only if the automated situation turns unsafe [5,6] by 2050 [7]. Thus, the near-future traffic would be a mixture of human-driven vehicles (HDVs) and MVs with various levels of sensing capabilities, which is called mixed traffic hereafter. Therefore, it is necessary to explore the possibility of using an MV equipped with a low-cost and popular sensor (e.g., a monocular camera), with the purpose of enhancing generalizability in mixed traffic to collect traffic data of the observed vehicles and feed them into the ITMSs.

In our previous studies [8,9], we have investigated the feasibility of using a vehicle equipped with a low-cost front-facing monocular camera with a built-in global positioning system (GPS) receiver (hereafter, we call this vehicle an ego vehicle) to observe another vehicle (hereafter, we call this vehicle the target vehicle) and estimate its speed, distance, and the lane it is in. After studies [8,9], a follow-up research question has been raised about the use of an ego vehicle in estimating the geolocation of the target vehicle, as accessing the vehicle's geolocation plays a critical role in modeling the traffic scene and making smart decisions by ITMSs.

Therefore, in this paper, we go beyond the lane-level target-vehicle localization presented in [9] and find the latitude and longitude of a target vehicle in a GPS coordinate system dynamically while both the ego vehicle and the target vehicle are moving in a metropolitan area. Although some research has been carried out

\* Corresponding author.

on utilizing an ego vehicle as a mobile sensor to estimate traffic data of the target vehicle, there is still very little scientific understanding of estimating the geolocation of HDVs based on ego-vehicle self-localization, image-based estimated distance to the target vehicle, and the relative angle between them by using a monocular camera with a built-in GPS receiver mounted on a mobile ego vehicle.

Therefore, the objective of this paper is to investigate the feasibility of using data from low-cost sensors (i.e., a monocular camera with a built-in GPS receiver) mounted on an ego vehicle to estimate the geolocation of a moving target vehicle. Our research question is defined as follows:

- RQ: How can the geolocation of a mobile target vehicle be dynamically estimated in a GPS coordinate system based on the vision of a front-facing low-cost monocular camera with a built-in GPS receiver on a mobile ego vehicle?

To address this research question, we proposed two approaches based on (1) object detection and image processing and (2) geometric computation by considering the camera's pitch angle and height from the road surface. In this regard, we extended the proposed algorithms presented in [9] by including the estimation of the distance and angle between the ego vehicle and the target vehicle.

To evaluate our proposed approaches and develop algorithms, we ran empirical experiments using real traffic data from a metropolitan area in Chengdu, China. We analyzed the findings by plotting the estimated target vehicle's trajectory on Google Maps and compared it with the ground-truth trajectory of the target vehicle. Additionally, the vector distance was used to quantitatively analyze the deviations between the estimated and the ground-truth geolocations of the target vehicle. The evaluation results confirmed that both approaches could estimate the geolocation of the target vehicles accurately.

The rest of the paper is organized as follows. Section 2 gives a brief overview of the recent history related to vehicle localization approaches. Section 3 explains the research strategy and methodology we propose. Section 4 presents the experiments and results of our approaches. The discussion is presented in Section 5. The last section concludes and proposes future studies.

## 2. Related work

To estimate the target vehicle's geolocation in a GPS coordinate system, we need to know the ego vehicle's geolocation and the target vehicle's location (the distance and angle between the ego vehicle and the target vehicle) [10]. This section presents a related work of these aspects briefly.

### 2.1. Ego-vehicle geolocation

There has been an increasing amount of literature on estimating the geolocation of ego vehicles, which is usually called self-localization (e.g., [11]). A GPS receiver is one of the most popular sensors for localizing ego vehicles [12]. Standard GPS receivers in the market have an accuracy of about 10–15 meters in 95% of the time [13]. To minimize the GPS receiver's estimation error in ego-vehicle localization, map matching is applied widely [14]. Huang et al., [14] classified the map matching algorithms into four categories: geometric theory, topology, probability statistics, and advanced model.

### 2.2. Target vehicle's location estimation

To date, various studies have investigated target vehicle's location estimation via monocular cameras regarding driving

safety measures, assistance, and autonomous navigation. For instance, Ifthekhar et al., [15] introduced an optical camera communications (OCC)-based cooperative vehicle positioning (CVP) technique. They proposed two approaches: (1) a neural network-based approach and (2) a computer vision-based approach to estimate the target vehicle's location. They considered two vehicles, one as an observing vehicle equipped with front-left and front-right cameras. Another vehicle was treated as a target vehicle, and its positioning was estimated based on its rear light-emitting diodes (LEDs). Simulation results showed that the accuracy achieved by the proposed neural network-based method was higher than the computer vision-based method [15]. Hayakawa et al., in [16] proposed a new approach based on integrating three deep neural networks to estimate the ego-motion and the target vehicle's state (e.g., 3D vehicle bounding box, depth, and optical flow). The experimental evaluations demonstrated that the distance error in the lateral and longitudinal directions were 1.19 m and 1.70 m, respectively.

Lee [17] focused on inter-vehicle distance estimate based on lane width. The proposed technique had a distance estimate error of less than 7%. Huang et al. [18] proposed a novel approach to estimate the inter-vehicle distance based on vanishing point detection, road segmentation, and vehicle detection. The ratio of true distance to image pixel was used to calculate the distance. In [18], a single-lens camera was utilized to capture data from urban/suburban roadways. Five image sequences of urban/suburban roads were utilized to verify the performance of the suggested method. The results showed average detection rate (DR), and false alarm rate (FAR) values of the approach are 82.21% and 16.16%, respectively [18]. Giesbrecht et al. [19] proposed a vision-based leader/follower system for an ego vehicle. The system was a combination of three main components: (1) a computer vision system for tracking the target vehicle based on color and the scale-invariant feature transform (SIFT), (2) a control system based on linear quadratic Gaussian control, and (3) a path following system. Their experiments showed that the mean and maximum error in the visual distance estimate were 0.72 m and 2.42 m, respectively, the follower speed was between 7.6 km/h and 10.2 km/h, and the follower separation was between 10.46 m and 23.71 m.

Taken together, although some research has been carried out on vehicle localization, more detailed empirical investigations are needed to dynamically estimate the target vehicle's geolocation in a GPS coordinate system via a monocular camera with the purpose of generating data to model the traffic scene and improve the ITMS performance.

## 3. Research strategy and methodology

In this paper, we proposed two new approaches by integrating deep learning, image processing, and geometric computation to use the vision sensing and self-localization capabilities of a mobile ego vehicle to estimate the geolocation of target vehicles. Fig. 1 illustrates our proposed research strategy. The components included in Fig. 1 are as follows:

### 3.1. Pre-processing

For the estimation of the target vehicle's geolocation, the geolocation of the ego vehicle is required. The collected latitude and longitude of the ego vehicle, which are usually collected by a GPS receiver, might be noisy. Collecting the geolocations of the ego vehicle accurately plays a vital role in accurately estimating the target vehicle's geolocations. Therefore, to enhance the accuracy of the ego-vehicle localization in the GPS coordinate system,

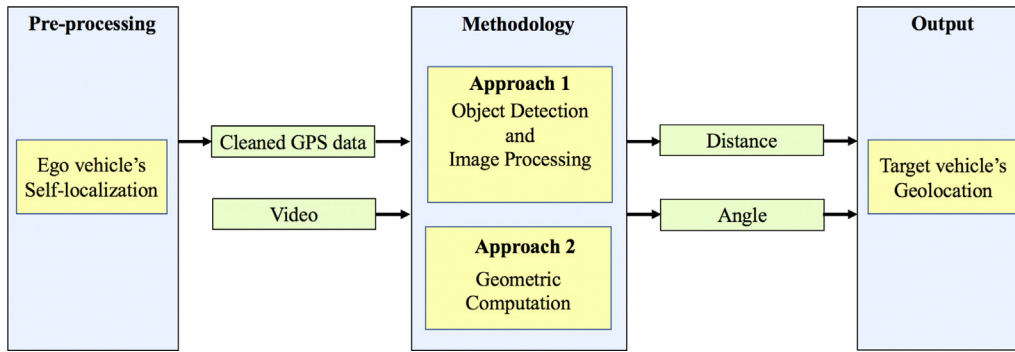


Fig. 1. The proposed steps in our research strategy.

the proposed approach in [20], based on cross-GPS validation, interpolation, best-fit, and map-matching techniques, is used.

### 3.2. Methodology

As shown in Fig. 2, to estimate the target vehicle's geolocation in the GPS coordinate system, in addition to the ego vehicle's geolocation, the distance  $d$  between the ego vehicle  $V_E$  and the target vehicle  $V_T$  and the clockwise angle  $\alpha$  between the north (N) and  $d$  are required [10].

In this regard, as Fig. 1 shows, we proposed two approaches, as follows:

#### 1. Approach 1: Object detection and image processing

- Estimating the distance  $d$

To begin the distance estimation process, we employed you only look once (YOLO)-v3 [21,22] to detect target vehicles via the ego vehicle's vision. YOLO-v3 is a well-documented open-source one-stage method for detecting objects. "YOLO-v3 is extremely fast and accurate" [23]. For example it is more than 1000x faster than R-CNN and 100x faster than Fast R-CNN [21,22]. Wang et al., [24] listed YOLO-v3 as the second most popular object detector models. Therefore, we chose to use YOLO-v3 in this study. We trained YOLO-v3 on the KITTI dataset [25], as this study focuses on traffic objects, and KITTI includes eight categories of traffic objects: car, van, truck, pedestrian, person\_sitting, cyclist, tram, and misc [25]. To estimate the distance  $d$  from the ego vehicle to the target vehicle, we followed the

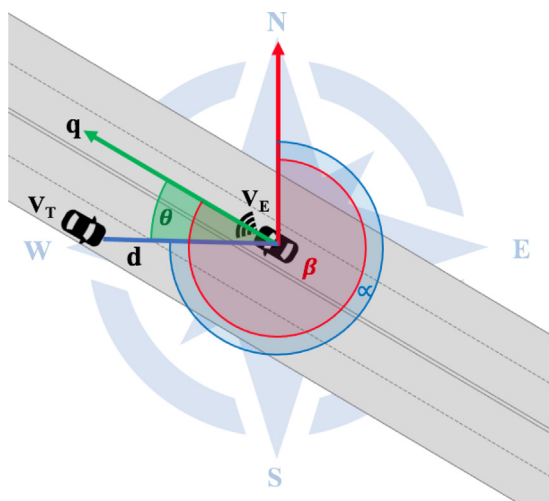


Fig. 2. The required parameters for estimating the target vehicle's geolocation.

approach proposed by Namazi et al. [8]. This approach [8] was based on the pinhole camera model by considering the real, pre-known size of the vehicle and the size of the bounding box added by YOLO-v3 around the target vehicle on the image plane. Distance  $d$  was calculated based on the average of the computed distances for both vehicle width and vehicle height by using a weight factor (i.e., 85% of the height and 15% of the width).

- Estimating the angle  $\alpha$

As presented in Fig. 2, in order to estimate the clockwise angle  $\alpha$  between the north N and  $d$ , we need to estimate angle  $\beta$ , which is the angle between the north N and the ego vehicle  $V_E$ 's movement direction  $q$ , as well as angle  $\theta$ , which is the angle between  $d$  and  $q$ .

- (a) Estimating the angle  $\beta$

To estimate the angle  $\beta$ , we need to identify the movement direction of the ego vehicle  $V_E$  based on its collected GPS coordinates in sequential frames as a start point  $(\phi_1, \lambda_1)$  and an end-point  $(\phi_2, \lambda_2)$  for all frames. We used Eq. 1 - Eq. 3 [10] to estimate angle  $\beta$  along the whole trajectory dynamically.

$$M = \sin(\lambda_2 - \lambda_1) \cdot \cos \phi_2 \tag{1}$$

$$N = \cos \phi_1 \cdot \sin \phi_2 - \sin \phi_1 \cdot \cos \phi_2 \cdot \cos(\lambda_2 - \lambda_1) \tag{2}$$

$$\beta = \text{atan2}(M, N) \tag{3}$$

- (b) Estimating the angle  $\theta$

The idea of estimating angle  $\theta$  in our first approach is presented in Fig. 3. In Fig. 3, the blue bounding box shows the target vehicle  $V_T$ . P is the central point on the bottom edge of the bounding box around the target vehicle  $V_T$ , and H is the central point of the image. Angle  $\theta$  is estimated based on the horizontal angle per pixel ( $\gamma$ ) in degrees and on the number of horizontal pixels between P and the vertical line passing

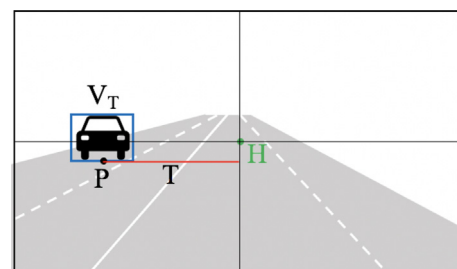


Fig. 3. The parameters used to estimate angle  $\theta$  between the ego vehicle and the target vehicle used in Approach 1.

through  $H$ , shown by a red line  $T$ .  $\gamma$  is estimated based on the camera's horizontal field of view (FOV) and the video's resolution. In this study, the video's resolution was  $960 \times 720$  pixels, and the camera's horizontal FOV was 86.7 degrees [26]. Therefore,  $\gamma$  is equal to 0.09 degrees. Angle  $\theta$  in degrees is estimated as follows:

$$\theta = T \cdot \gamma \quad (4)$$

(c) Estimating the angle  $\alpha$

To estimate the angle  $\alpha$ , we considered three different conditions, as follows:

- If the target vehicle drives in the same lane as the ego-vehicle, then  $\alpha = \beta$  and  $\theta = 0$ .
- If the target vehicle drives on the left side of the ego vehicle, then, as Fig. 4, (a) shows,  $\alpha = \beta - \theta$ .
- If the target vehicle drives on the right side of the ego vehicle, then, as Fig. 4, (b) shows,  $\alpha = \beta + \theta$ .

2. Approach 2: Geometric computation

The main idea of this approach is to transform 2D pixel coordinates of point  $P$  into 3D world coordinates. By assessing the 3D world coordinates of point  $P$ , we would be able to estimate distance  $d$  and angle  $\alpha$ , which are needed to estimate the target vehicle's geolocation.

In this regard, we utilized a pinhole camera model as shown in Fig. 5. In this figure,  $C$  is the perspective center of the camera and the origin of the camera coordinate frame (CCF). Three unit vectors of the CCF are represented by  $y_1, y_2$ , and  $y_3$ . The image coordinate frame (ICF) is centered at principal point  $H$  with unit vectors  $r_1$  and  $r_2$ . The principal axis passes through  $C$  and  $H$  and is perpendicular to the image plane. The distance from  $C$  to the image plane is  $f$ , which is the camera's focal length. The image plane carries a 2D pixel coordinate frame (PCF) with unit vectors  $z_1$  and  $z_2$ . The image plane is subdivided into  $n_h$  pixels horizontally and  $n_v$  pixels vertically. To project the detected vehicle on the image onto the real world, we need to transform point  $P$  with pixel coordinates  $(p_1, p_2)$  (which is the central point on the

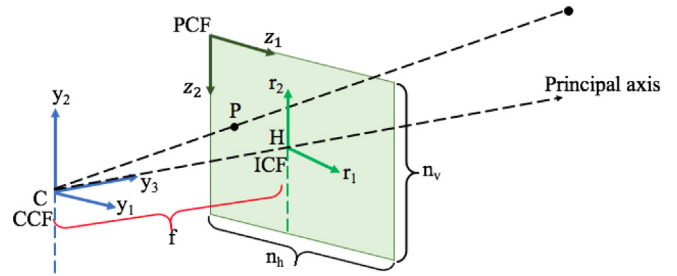


Fig. 5. The pinhole camera model used in Approach 2.

button edge of the bounding box around the target vehicle) into a 3D camera coordinate representation  $(w_1, w_2, w_3)$ .

In this regard, we first need to identify the 3D coordinates of point  $P$  in 3D camera coordinates. Based on Fig. 6, the 3D coordinates of point  $P$  in 3D camera coordinates are presented in Eq. 5.

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} p_1 - h_1 \\ -(p_2 - h_2) \\ f \end{pmatrix} \quad (5)$$

In the follow-up step, we need to identify the 3D coordinates of point  $P$  in 3D world coordinates. In this step, we temporally assumed that the camera's pitch angle, yaw angle, and roll angle were equal to 0. The height of the camera mounted on the ego vehicle from the road surface is named  $h$ , and we assumed that the world coordinates are located on the road surface. The 3D coordinates of point  $P$  in 3D world coordinates  $(w_1, w_2, w_3)$  are given in Eq. 6.

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} p_1 - h_1 \\ -(p_2 - h_2) + h \\ f \end{pmatrix} \quad (6)$$

After that, we need to find a mathematical expression for all points that lie on the viewing ray from camera center  $C$  through point  $P$  in world coordinates, as presented in Eq. 7, where  $\xi$  defines any position along the viewing ray.

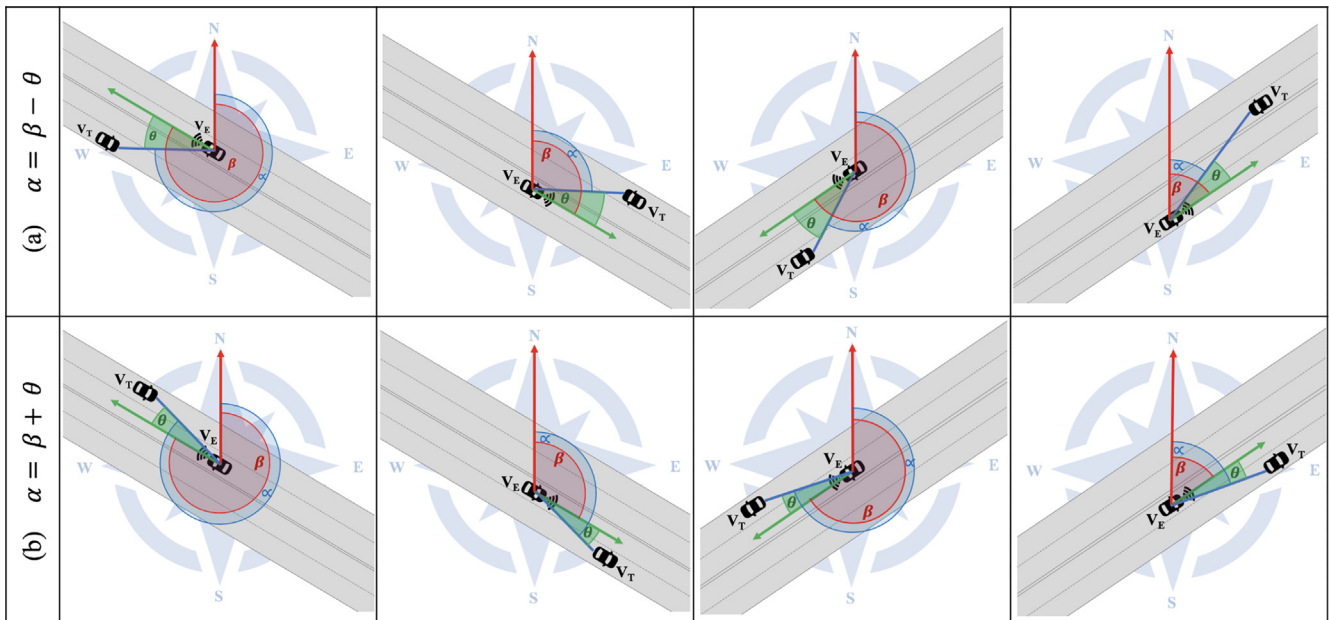


Fig. 4. The mathematical relations between  $\alpha$ ,  $\beta$ , and  $\theta$ .

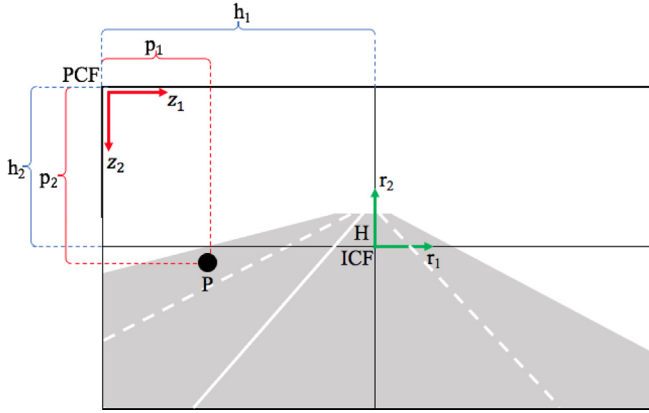


Fig. 6. The image plane used in Approach 2.

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} 0 \\ h \\ 0 \end{pmatrix} + \xi \cdot \begin{pmatrix} p_1 - h_1 \\ -(p_2 - h_2) \\ f \end{pmatrix} \quad (7)$$

As we assumed, the world coordinate system is located on the road surface; therefore, the height of any point on the road surface in the world coordinate system is equal to 0. So we can express Eq. 7 as Eq. 8.

$$\begin{pmatrix} w_1 \\ 0 \\ w_3 \end{pmatrix} = \begin{pmatrix} 0 \\ h \\ 0 \end{pmatrix} + \xi \cdot \begin{pmatrix} p_1 - h_1 \\ -(p_2 - h_2) \\ f \end{pmatrix} \quad (8)$$

Finally, to increase the accuracy of this estimation, we need to consider the camera's pitch angle  $\sigma$ . Rotating the camera by angle  $\sigma$  has no effect on the location of point P in the  $w_1$  direction, but the point's location in the  $w_2$  and  $w_3$  directions will be affected by this rotation. Therefore, the following rotation is applied:

$$\begin{pmatrix} w_1 \\ 0 \\ w_3 \end{pmatrix} = \begin{pmatrix} 0 \\ h \\ 0 \end{pmatrix} + \xi \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(-\sigma) & -\sin(-\sigma) \\ 0 & \sin(-\sigma) & \cos(-\sigma) \end{pmatrix} \cdot \begin{pmatrix} p_1 - h_1 \\ -(p_2 - h_2) \\ f \end{pmatrix} \quad (9)$$

Therefore,  $\xi$ ,  $w_1$ , and  $w_3$  are calculated as follows:

$$\xi = \frac{-h}{(\cos(-\sigma) \cdot (-p_2 + h_2) - \sin(-\sigma) \cdot f)} \quad (10)$$

$$w_1 = \xi \cdot (p_1 - h_1) \quad (11)$$

$$w_3 = \xi \cdot (\sin(-\sigma) \cdot (-p_2 + h_2) + \cos(-\sigma) \cdot f) \quad (12)$$

To calculate  $\xi$ ,  $w_1$ , and  $w_3$ , we need to estimate the camera's height from the road surface  $h$ , the camera's pitch angle  $\sigma$ , and the camera's focal length  $f$ .

- Estimating the camera's focal length  $f$

In Approach 2, the camera's focal length  $f$  in pixels is calculated based on the trigonometric relation presented in Eq. (13). We used the horizontal number of pixels  $n_h$  from the video's resolution and the camera's horizontal FOV  $\rho$  in degrees [26].

$$f = \frac{n_h}{2 \cdot \tan(\frac{\rho}{2})} \quad (13)$$

- Estimating the camera's pitch angle  $\sigma$

To estimate the camera's pitch angle  $\sigma$ , we used a vanishing point estimated based on the lane detection.

To detect lanes, as we presented in [8,9], we used canny edge detection [27] and the progressive probabilistic Hough transform [28,29]. In this study, we go further to identify the vanishing point based on the detected parallel lines on the road nearby the ego vehicle.

To estimate the pitch angle  $\sigma$ , we used the camera's focal length ( $f$ ) and the vertical differences between the principal point  $H = (h_1, h_2)$  and the vanishing point  $J = (j_1, j_2)$ , as shown in Fig. 7. In this figure, the blue lines represent the detected parallel lines on the road nearby the ego vehicle in a perspective view. Based on this figure, the camera's pitch angle  $\sigma$  can be calculated by Eq. 14.

$$\sigma = \text{atan2}(j_2 - h_2, f) \quad (14)$$

- Estimating the camera's height  $h$  from the road surface

To estimate the height  $h$  of the camera mounted on the ego vehicle from the road surface by considering the camera's pitch angle  $\sigma$ , we applied Thales's theorem [30]. Thales's theorem in this context is presented in Fig. 8. The variables in this figure are defined as follows:

$$A = \frac{f}{\cos \sigma} \quad (15)$$

$$B = h \cdot \tan \sigma \quad (16)$$

$$E = \frac{h}{\cos \sigma} \quad (17)$$

$$G = f \cdot \tan \sigma \quad (18)$$

$$K = G + N \quad (19)$$

To estimate the camera's height  $h$  based on Thales's theorem and the estimated distance  $d$  by Approach 1, we have the following equation.

$$\frac{K}{E} = \frac{A}{d + B} \quad (20)$$

By simplifying Eq. 20,  $h$  in meters is calculated as follows:

$$h = \frac{(N + f \cdot \tan \sigma) \cdot d \cdot (\cos \sigma)^2}{f - (N + f \cdot \tan \sigma) \cdot \tan \sigma \cdot (\cos \sigma)^2} \quad (21)$$

Finally, by estimating the camera's height from the road surface  $h$ , the camera's pitch angle  $\sigma$ , and the camera's focal length  $f$ , we can calculate  $\xi$ ,  $w_1$ , and  $w_3$ . Because  $w_1$  and  $w_3$  represent point P

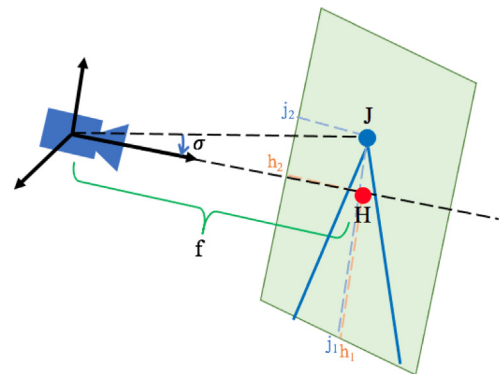


Fig. 7. The camera's pitch angle  $\sigma$  and vanishing point used in Approach 2.

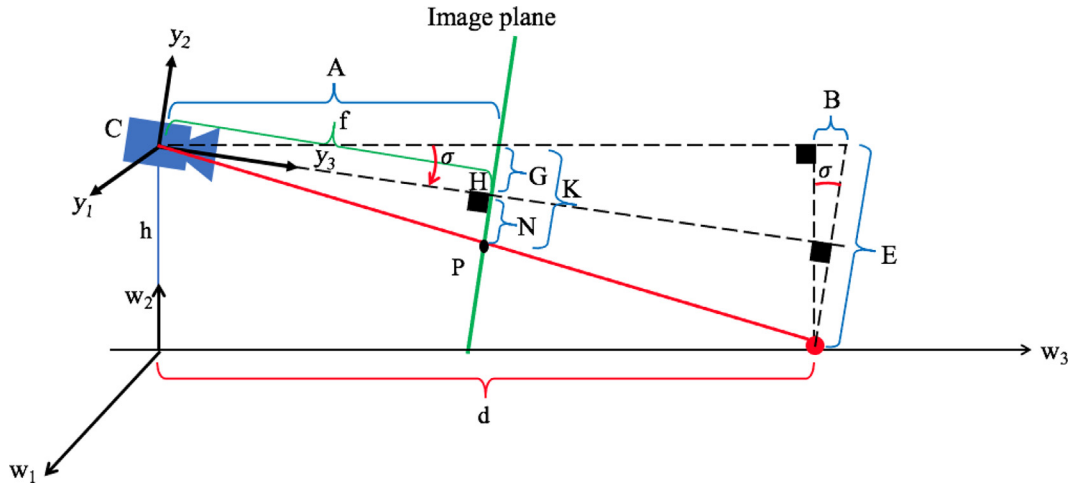


Fig. 8. The camera's height from the road surface by considering the camera's pitch angle  $\sigma$  used in Approach 2.

in 3D world coordinates, where  $w_2=0$ , we can estimate distance  $d$  by Approach 2 based on the Euclidean distance [31] between  $w_1$  and  $w_3$  as presented in Eq. 22, and estimate angle  $\theta$  based on trigonometry [32] presented in Eq. 23. Finally, angle  $\alpha$  can be estimated based on the proposed conditions in Section 3.2.1.

$$d = \sqrt{w_1^2 + w_3^2} \quad (22)$$

$$\theta = \text{atan2}(w_3, w_1) \quad (23)$$

### 3.3. Estimating the geolocation of the target vehicle

To estimate the target vehicle's geolocation with both approaches, we used Eq. 24 - Eq. 27 [10]. In these formulas, the variables are as below:  $(\ell_1, g_1)$  represent the geolocation of the ego vehicle  $(\ell_2, g_2)$  represent the geolocation of the target vehicle  $R$  represents the Earth's radius.

$d$  represents the estimated distance between the ego vehicle and the target vehicle by both approaches

$\alpha$  represents the estimated angle between the north  $N$  and  $d$  by both approaches

$$\ell_2 = \text{asin}(\sin(\ell_1) \cdot \cos(d/R) + \cos(\ell_1) \cdot \sin(d/R) \cdot \cos(\alpha)) \quad (24)$$

$$U = \sin(\alpha) \cdot \sin(d/R) \cdot \cos(\ell_1) \quad (25)$$

$$V = \cos(d/R) - \sin(\ell_1) \cdot \sin(\ell_2) \quad (26)$$

$$g_2 = g_1 + \text{atan2}(U, V) \quad (27)$$

## 4. Experiments and results

In this paper, we carried out experiments using real-world traffic data to demonstrate the effectiveness of both proposed approaches for estimating the target vehicle's geolocation by an ego vehicle's vision.

### 4.1. Data collection

We used three vehicles and drove them by the following the pre-defined scenarios in Chengdu, China. All vehicles were equipped with two GoPro Hero 7 cameras, and each camera included a built-in GPS receiver. We used the GoPro Hero 7 camera as it provides us with both visual information and GPS data of the

vehicles. One of the cameras mounted on the front window glass looked forward through the window, and another mounted on the front window glass looked backward. The purpose of mounting two cameras on the same vehicle was to improve the ego vehicle's self-location in the pre-processing step [20], and to collect more data for future studies. To estimate the target's geolocation, this study used only the footage collected from the camera mounted on the front window glass. We used one of these vehicles as an ego vehicle. The other two vehicles were treated as target vehicles. The GPS data collected by the target vehicles were used as ground truth to assess our proposed approaches.

The settings of the GoPro Hero 7 cameras we used were as follows. The used mode was  $4 \times 3$  and linear with a zoom = 0%. During the recording of the footage, the video's resolution and the frame rate were  $1920 \times 1440$  and 60 frames per second (FPS), respectively. We adjusted the video's resolution to  $960 \times 720$  and the frame rate to 1 FPS to apply pre-processing and vehicle detection.

Fig. 9 shows the studied scenarios, called Scenario S1 and Scenario S2. In Scenario S1, ego vehicle  $v_3$  and target vehicles  $v_1$  and  $v_2$  are driven in the same direction on a straight trajectory. The purpose of this scenario was to evaluate our proposed approaches with one of the target vehicles driving on the same lane as the ego vehicle and the other target vehicle driving on the next lane. In Scenario S2, ego vehicle  $v_3$  and target vehicles  $v_1$  and  $v_2$  are driven in opposite directions on a straight road. In the scenario in which the vehicles are driven in opposite directions, the period between detecting a target vehicle via an ego vehicle

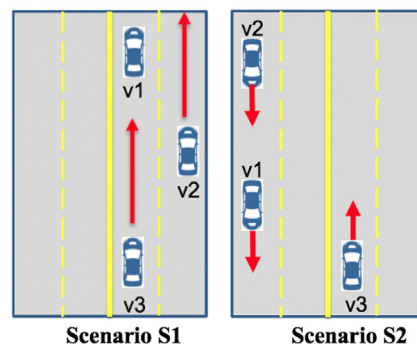


Fig. 9. The scenarios studied using both approaches.

until both vehicles pass each other is short, so the number of estimated locations is limited to this short period.

### 4.2. Experiments

The experiments were run using a laptop with a 3.1 GHz Intel Core i5 processor and Intel Iris Plus Graphics 650 1536 MB. As we explained in Section 4.1, we chose to use 1 FPS when analyzing the video, with the purpose of making a trade-off between the amount of generated data and the running time of the system. Also, as the vehicles' speeds were low (between 19.55 km/h and 30.18 km/h on average), there were hardly any informative changes in the vehicle's speed, distance, angle, and location within less than one second. Therefore, analyzing the data with a higher frequency could not provide much extra information. We measured the running time based on the experimental studies using Approaches 1 and 2 after the data pre-processing. We found that the system based on Approach 1 ran 0.680 FPS on average, and that the system based on Approach 2 ran 0.684 FPS on average, to output the geolocation of the target vehicle from the input videos and the pre-processed ego vehicle's geolocations.

#### 4.2.1. Evaluation

The evaluation is done in two steps: (1) plotting the estimated geolocations (i.e., latitude and longitude) of the target vehicles on Google Maps and (2) analyzing the distance vector between the estimated target vehicle's geolocations by both proposed approaches and the ground-truth data.

Fig. 10 and Fig. 11 present the outputs of the experiments related to Scenario S1 and Scenario S2, respectively. In these figures, the white polyline in (a) and (d) shows the ground-truth trajectory of the observed target vehicle. The red polyline in (b) and (e) shows the trajectory of the target vehicle estimated with Approach 1. The blue polyline in (c) and (f) shows the trajectory of the target vehicle estimated with Approach 2. These figures show that the trajectories of the target vehicle estimated with both approaches are plotted on the correct lane of the road and that they almost overlap with the ground-truth trajectory. This means that both approaches enable us to estimate the trajectory of the target vehicle accurately on the right lane of the road. As expected, with Scenario S2, the number of plotted positions along the trajectory, presented in Fig. 11, are limited (2–3 points) because of the opposite movement directions of the ego vehicle and target vehicle and the short sensing time.

To analyze the estimated geolocations of the target vehicle numerically, we used the distance vector between the ground truth and the geolocations estimated by both approaches. This analysis provides information regarding the estimation deviation in our proposed approaches. Fig. 12 visualizes our findings related to Scenario S1 and Scenario S2. Our numerical findings are summarized in Table 1, as well. As Fig. 12 and Table 1 show, the geolocation estimation deviation (based on the absolute values) with Approach 1 is on average between 1.38 m and 3.54 m. The geolocation estimation deviation with Approach 2 is on average between 1.4 m and 3.51 m. Fig. 12 (a) and (b) shows a slightly upward trend between the plotted points. This result may be explained by the fact that the collected data by a GPS receiver to provide the ego vehicle's location and ground truth data of the target vehicle's position were not noise-free. As we expected, Fig. 12 (e) - (h) represent the limited points as Scenario S2 focused on studying the vehicle movement in the opposite directions and the sensing lifetime was limited. In addition, as Table 1 shows, the highest on average geolocation estimation deviation with both approaches is obtained in Scenario S2, when v3 observes v1. A possible explanation for this might be that as in this scenario, only limited geolocations (2–3) were estimated, so the estimation deviation of one point has a big effect on the average error.

To analyze our proposed approaches further, we applied the root mean square error (RMSE) to the distance vector between the estimated geolocations and the ground truth to show the estimation deviation. The calculated RMSE related to Approach 1 was between 1.5 m and 3.7 m (2.39 m on average). The calculated RMSE related to Approach 2 was between 1.55 m and 3.63 m (2.37 m on average). Overall, these results indicate that, in the studied scenarios, Approach 2 is slightly (about 0.02 m on average) better than Approach 1.

As the speed of the vehicles and the distance between them may influence our estimation accuracy, we studied the effect of the speed of the ego vehicle and target vehicle and the distance between them on estimating the target vehicle's geolocation. As an example, we presented our findings related to Scenario S1 in Fig. 13. Fig. 13 shows that in the case in which v3 observes v1, with both vehicles driving on the same lane, changes in the distance and speed have no significant effect on our estimation accuracy. However, when v3 observes v2, with both vehicles driving on different lanes, increasing the distance between the vehicles, as caused by changes in the vehicles' speed, increases the estimation deviation. However, as the speed of the vehicles and the distance between

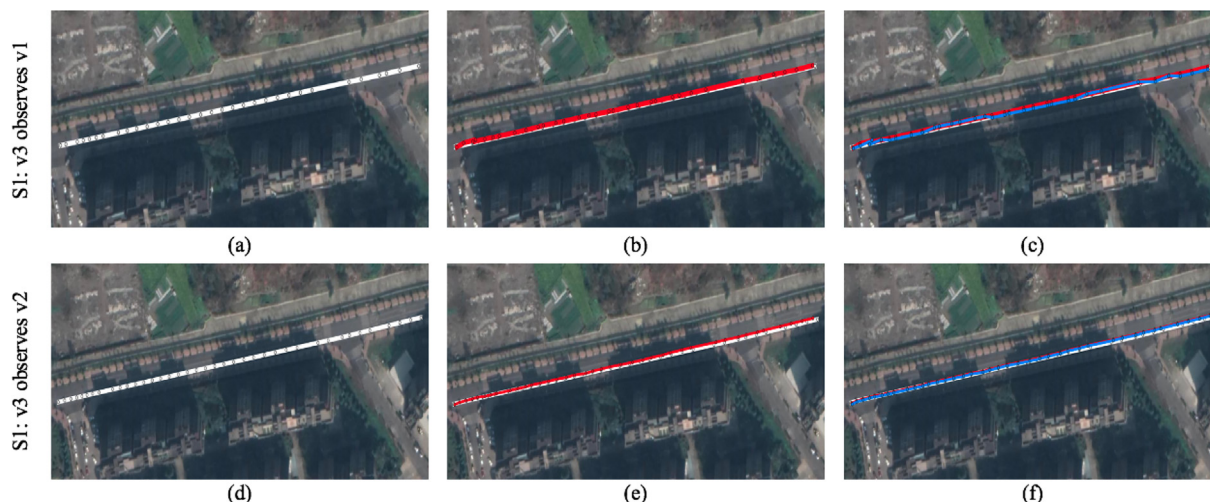
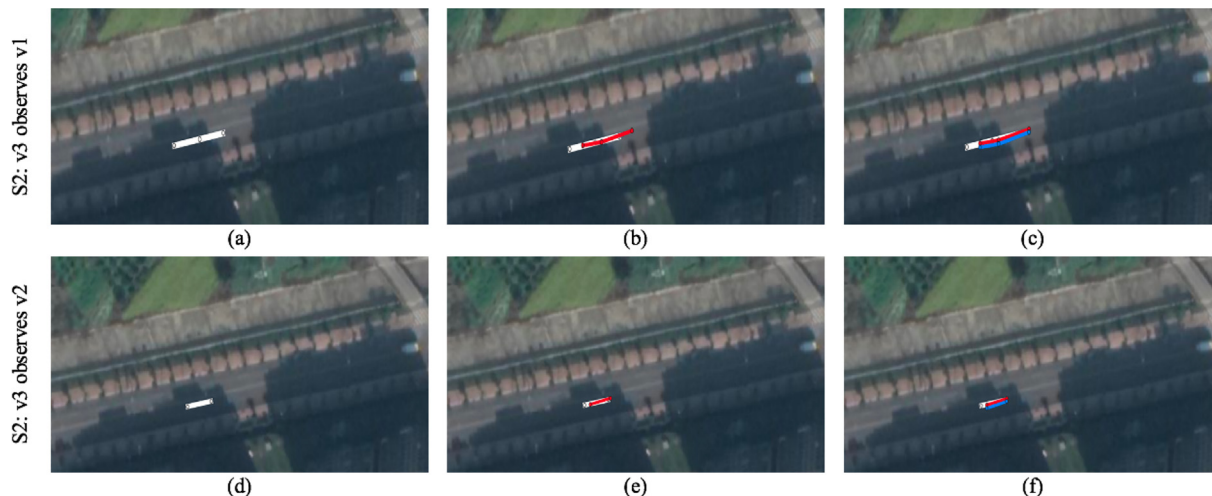
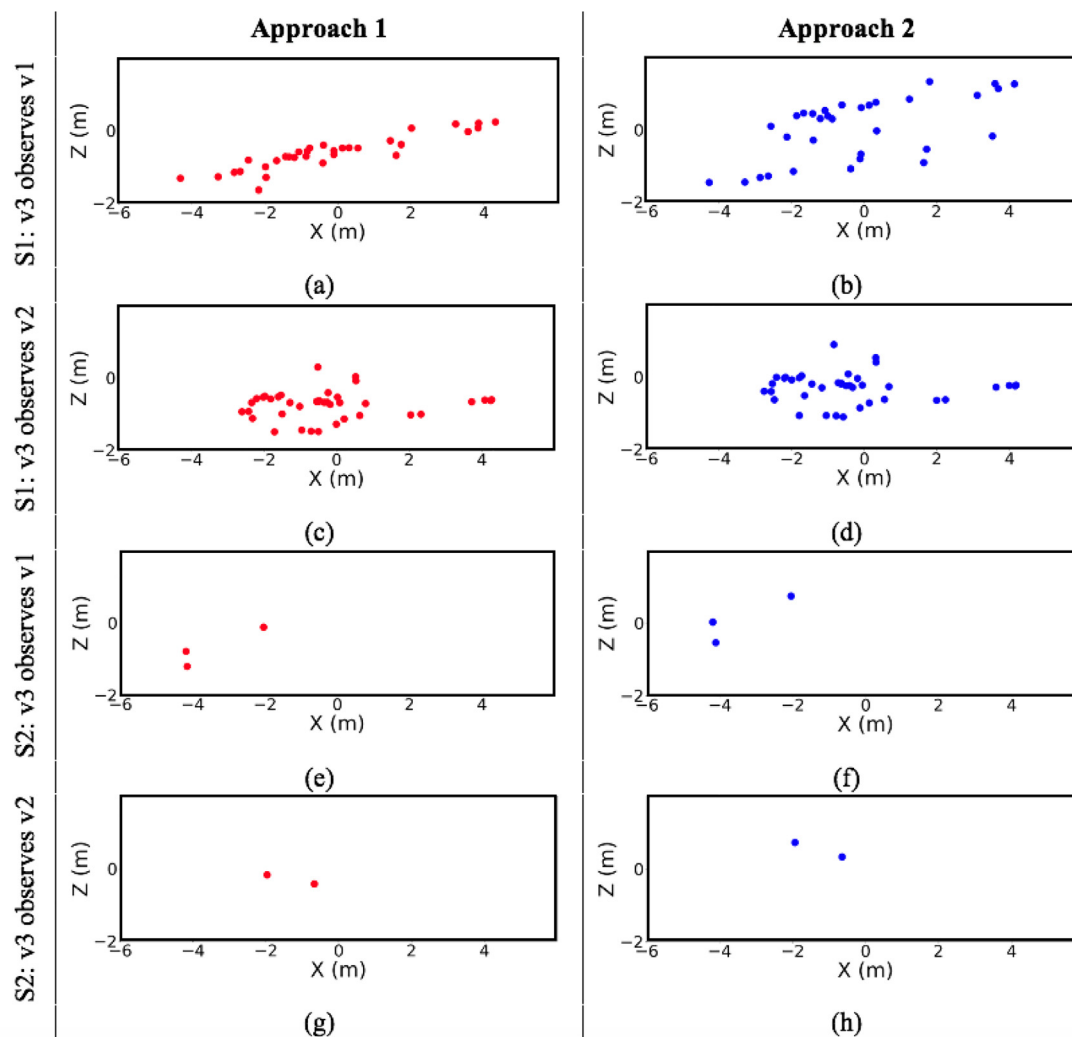


Fig. 10. The estimated trajectory with Scenario S1 on the map. Two cases are considered: (1) ego vehicle v3 observes target vehicle v1 (a-c) and (2) ego vehicle v3 observes target vehicle v2 (d-f).



**Fig. 11.** The estimated trajectory with Scenario S2 on the map. Two cases are considered: (1) ego vehicle v3 observes target vehicle v1 (a-c) and (2) ego vehicle v3 observes target vehicle v2 (d-f).

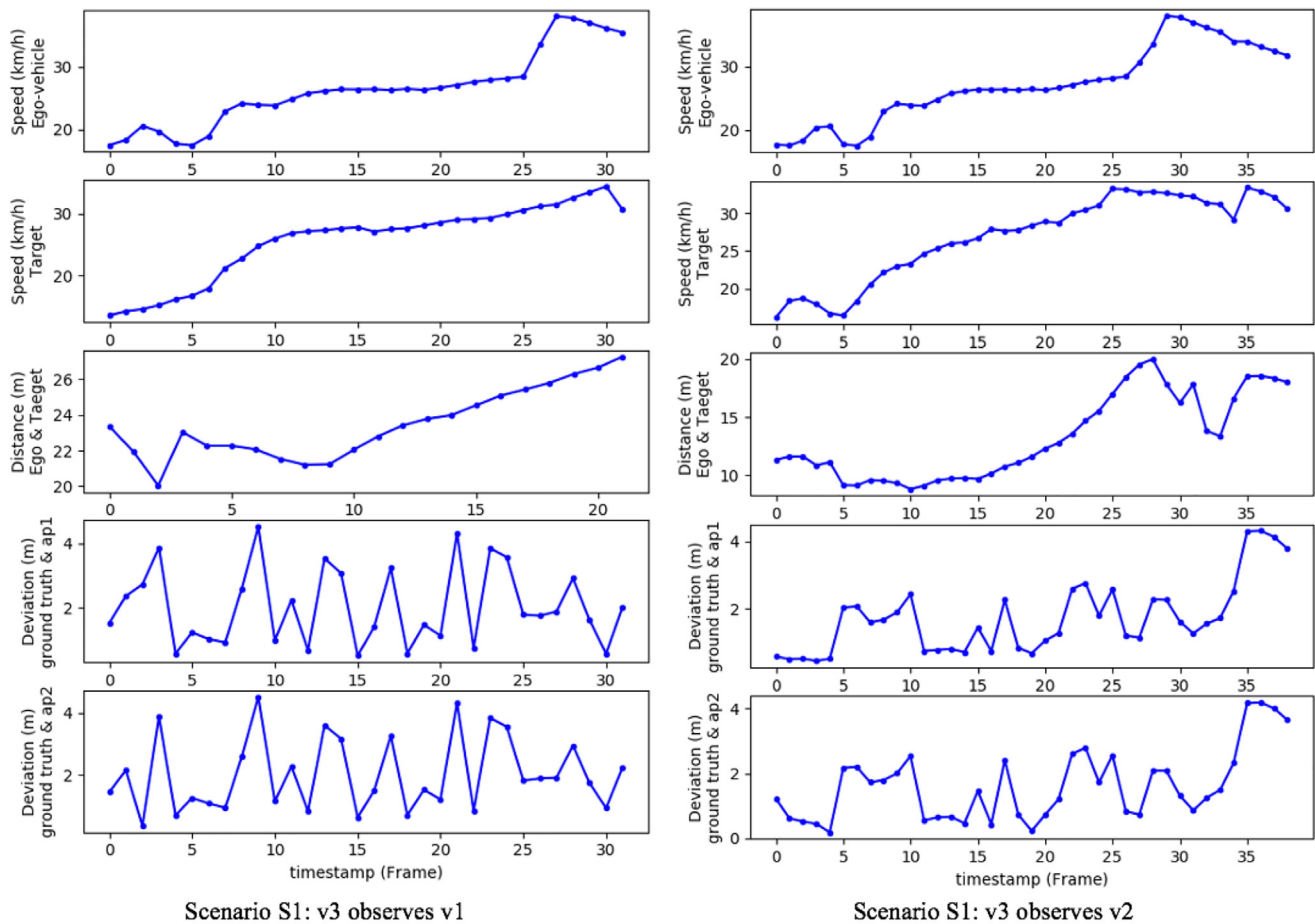


**Fig. 12.** The distance vectors between ground truth and estimated geolocations with Approach 1 (a, c, e, and g) and Approach 2 (b, d, f, and h) for Scenario S1 and Scenario S2. Two cases are considered: (1) ego-vehicle v3 observes target vehicle v1, and (2) ego-vehicle v3 observes target vehicle v2. X shows longitudinal and Z shows lateral directions.



**Table 1**  
Evaluation results.

S#	V#	Estimation deviation of Approach 1 (m)				Estimation deviation of Approach 2 (m)			
		Min	Avg	Max	RMSE	Min	Avg	Max	RMSE
S1	v3 observes v1	0.50	2.03	4.51	2.35	0.35	2.02	4.51	2.34
	v3 observes v2	0.47	1.74	4.31	2.03	0.19	1.63	4.18	1.96
S2	v3 observes v1	2.040	3.54	4.33	3.70	2.18	3.51	4.20	3.63
	v3 observes v2	0.79	1.38	1.97	1.50	0.72	1.4	2.07	1.55



**Fig. 13.** Plotting the deviations with regard to ground truth to study the effect of speed and distance (here shown for Scenario S1.).

**Table 2**  
Traffic data measurements with Scenario S1 and Scenario S2.

S#	V#	Ego vehicle speed (km/h)			Target vehicle speed (km/h)			Dist. ego and target vehicles (m)		
		Min	Avg	Max	Min	Avg	Max	Min	Avg	Max
S1	v3, v1	17.50	26.38	38.04	13.64	25.65	34.35	20.04	25.02	38.84
	v3, v2	17.50	27.14	38.04	16.19	26.94	33.4	8.81	13.25	20.01
S2	v3, v1	16.98	19.55	22.04	29.12	29.40	29.55	17.35	30.78	44.14
	v3, v2	22.04	23.2	24.36	29.63	30.18	30.74	20.53	27.95	35.37

them were limited in the studied scenarios, more studies are needed to validate these findings in the future. The extracted traffic data (i.e., the ego vehicle's and the target vehicle's speed and the distance between the vehicles) are summarized in [Table 2](#).

#### 4.2.2. Comparison of Approach 1 with Approach 2

Although Approach 2 can estimate the target vehicle's geolocation slightly better than Approach 1 on average, it cannot always be the optimum approach. Therefore, we investigated the devia-

tions between the ground truth and estimated geolocations in the longitudinal and lateral directions in time series using both approaches. Our findings related to Scenario S1 as an example are presented in Fig. 14.

As Fig. 14 shows, the deviations between the ground truth and the estimated geolocations in the longitudinal direction (a) with Approach 1 and Approach 2 are almost overlapped. However, this parameter in the lateral direction (b) is more different. This result may be explained by the fact that the ego vehicle's geolocation was utilized in estimating the angle between the ego vehicle and the target vehicle. However, the used ego vehicle's geolocation is not noise-free. In addition, the ground truth data to apply the estimation deviation were collected by such GPS receivers, as well. Another possible explanation for this is that the employed methodology to estimate the distance between ego vehicle and target vehicle with both approaches were different. The first approach relied on the accuracy of the bounding box added by YOLO-v3 around the target vehicle, and the second approach used the central point on the bottom edge of the bounding box. In general, by considering the proposed methodology in each approach, we can conclude in case the lane marks and the vanishing point, which are needed by the second approach, are available, we can use Approach 2 as it is identified as the more accurate approach; otherwise, Approach 1 can be used. In both cases, enhancing the accuracy of the GPS receiver accuracy and the vehicle detection algorithms are vital.

## 5. Discussion

To date, most of the existing studies on collecting traffic data focused on two main approaches: (1) stationary sensors and (2) V2V and V2I communications, which require most of the vehicles to have sensing and communication capabilities. However, the much-debated question is how to estimate the geolocation of the target vehicle in mixed traffic. To address this gap, this study proposed two approaches to dynamically estimating the target vehicle's geolocation based on an ego vehicle's vision capability.

### 5.1. Comparison with related work

As presented in Section 2, most of the existing studies regarding estimating the position of the target vehicle focused on the relative target vehicle's location estimation (e.g., [15,19]) and inter-vehicle distance estimation (e.g., [17,18]). Despite the importance of the target-vehicle localization, there remains a paucity of empirical studies on estimating the target vehicle's geolocation in a GPS coordinate system in order to enhance the ITMS awareness about the traffic scene. In addition, to be able to generalize the proposed

approaches in reality, using real traffic data to run experiments is vital; however, most of the studies applied the experiments by using a simulator (e.g., [15]). Moreover, studying the real traffic data from the urban area, which can reflect the possible estimation uncertainties and sources of noise (e.g., tall buildings that affect the GPS receiver accuracy) are important.

Therefore, the main objective of this study is to go beyond the target vehicle localization and estimate the relative angles besides estimating distance and ego vehicle's geolocation to be able to estimate the geolocation of the target vehicle dynamically. To estimate the ego vehicle's geolocation based on a low-cost GPS receiver, we used the proposed approach in [20]. To estimate distance and relative angle, we proposed two new approaches based on object detection and image processing and geometric computation. To assess our proposed approaches, we developed a new system and ran experiments on real traffic data collected from the urban area.

The experiments on estimating the target vehicle's geolocation showed that the estimation deviation with Approach 1 was on average between 1.38 m to 3.54 m. The results with Approach 2 were between 1.4 m and 3.51 m. In our study, the ego vehicle's speed was between 16.98 km/h and 38.04 km/h, the target vehicle's speed was between 13.64 km/h and 34.35 km/h, and the distance between the ego vehicle and the target vehicle varied between 8.81 m and 44.14 m. Comparison of our findings with those of similar studies focused on estimating the distance to the target vehicle (e.g., [19]) confirms that our geolocation estimation deviation is reasonable. For example, the approach proposed by Giesbrecht et al. [19] yielded an estimated distance with a mean and maximum error of 0.72 m and 2.42 m, respectively, with a follower speed between 7.6 km/h and 10.2 km/h and a follower separation between 10.46 m and 23.71 m. However, our experiments were applied to real traffic data collected from the urban area by considering the higher speed and distance.

### 5.2. Limitations of our proposed approaches

Although the experiments confirmed that both our proposed approaches were able to estimate the geolocation of the target vehicle accurately on the right lane of the road, each approach has some pros and cons. For instance, Approach 1 is tightly connected to the vehicle detection accuracy because we employed the size of the bounding box added by YOLO around the target vehicle to estimate its distance. However, during the experiments, we observed that these bounding boxes were shaking during the trajectory, and the size of them varies between frames, which can affect the accuracy of the distance estimation.

Approach 2 has some limitations as well. For instance, like Approach 1, Approach 2 is tightly dependent on the accuracy of

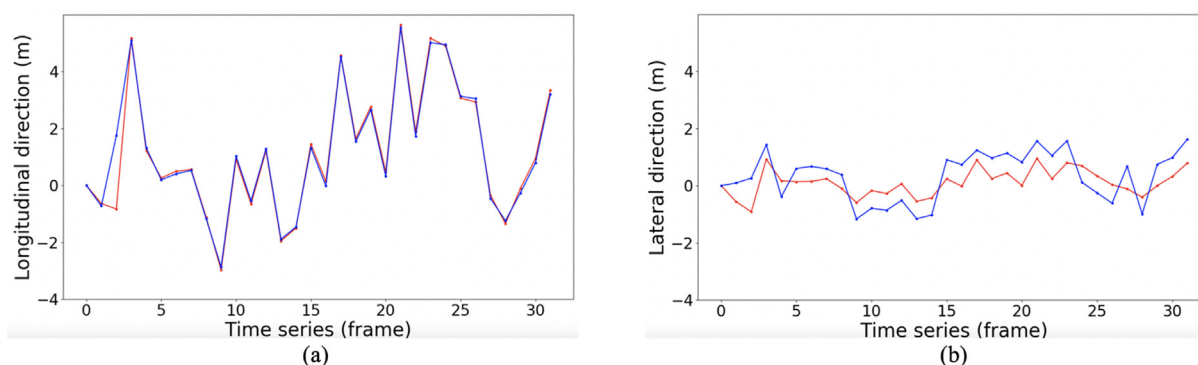


Fig. 14. The deviation between the ground truth and the estimated geolocations with Approach 1 (the red polyline) and Approach 2 (the blue polyline) in the lateral and longitudinal directions in Scenario S1.

the central point on the bottom edge of the bounding box around the target vehicle, which is used to transform 2D space into a 3D space. Therefore, it has a direct effect on estimating the distance of the target vehicle. In addition, the camera's pitch angle was considered in Approach 2 to enhance the localization accuracy. To estimate the camera's pitch angle, we used a vanishing point based on the parallel lane on the road. Therefore, enhancing the lane detection accuracy would be beneficial for enhancing the accuracy. Furthermore, this pitch angle was caused by our manual installation of the camera with a suction cup on the vehicle; therefore, installing the camera more precisely is highly recommended.

In addition, in the both Approaches, to estimate the angle, we used the ego vehicle's movement direction and the central point on the bottom edge of the bounding box around the target vehicle. Therefore, the estimation error of each of these parameters has a negative effect on the angle estimation accuracy. As the ego vehicle's movement direction was estimated based on the ego vehicle's GPS data, if the GPS data is noisy, the estimation deviation will increase. The same is true for the central point on the bottom edge of the bounding box around the target vehicle, which is estimated by YOLO. The experiments showed that the central point was not stable and was shaking between frames. Therefore, to mitigate the estimation deviation, enhancing the accuracy of the YOLO and GPS data is needed.

Another source of uncertainty is that the noisy low-cost GPS receiver used to collect the ground truth data in a metropolitan area surrounded by many tall buildings. Although we applied data pre-processing to mitigate the GPS receiver noise, it is still not noise-free. Moreover, since the study was limited to vehicle movements along straight streets with limited scenarios, more studies are needed to be able to generalize our proposed approaches by considering various vehicle movements scenarios.

## 6. Conclusion

The main goal of the current study was to dynamically estimate the geolocation of mobile target vehicles via a low-cost front-facing monocular camera on a mobile ego vehicle. To estimate the target vehicle's geolocation, the distance between the ego vehicle and the target vehicle and the relative angle are needed. In this regard, we proposed two approaches: (1) object detection and image processing and (2) geometric computation.

The results of the evaluation using real traffic data confirmed that our algorithms were able to estimate the geolocation of the target vehicles accurately. Taken together, these findings confirmed the feasibility of a vehicle-mounted monocular camera for estimating the location of target vehicles in mixed traffic. The present study lays the groundwork for future research on using an ego vehicle as a mobile sensor to collect traffic data to reduce the traffic cost and improve ITS efficiency. Further studies that take these data types into account will be needed to increase the accuracy and enhance the generalizability by considering various scenarios.

## CRedit authorship contribution statement

**Elnaz Namazi:** Conceptualization, Methodology, Data curation, Visualization, Investigation, Writing - original draft, Validation, Writing - review & editing. **Rudolf Mester:** Conceptualization, Methodology, Supervision, Validation, Writing - review & editing. **Chauru Lu:** Conceptualization, Supervision, Validation. **Jingyue Li:** Conceptualization, Supervision, Validation, Writing - review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] E. Namazi, J. Li, C. Lu, Intelligent intersection management systems considering autonomous vehicles: A systematic literature review, *IEEE Access* 7 (2019) 91946–91965.
- [2] H.A. Oliveira, F.R. Barbosa, O.M. Almeida, A.P. Braga, A vehicle classification based on inductive loop detectors using artificial neural networks, in: 2010 9th IEEE/IAS International Conference on Industry Applications-INDUSCON 2010, IEEE, 2010, pp. 1–6.
- [3] H.A. Kurdi, Review of closed circuit television (cctv) techniques for vehicles traffic management, *International Journal of Computer Science & Information Technology* 6 (2) (2014) 199.
- [4] B.A. Kiisa Wiegand, Challenges of the day-today operation of a traffic monitoring program (2016).
- [5] M.N. Ahangar, Q.Z. Ahmed, F.A. Khan, M. Hafeez, A survey of autonomous vehicles: enabling communication technologies and challenges, *Sensors* 21 (3) (2021) 706.
- [6] Sae international releases updated visual chart for its 'levels of driving automation' standard for self-driving vehicles. <https://www.sae.org/news/press-room/2018/12/sae-international-releases-updated-visual-chart-for-its-%E2%80%99Clevels-of-driving-automation%E2%80%9D-standard-for-self-driving-vehicles..>
- [7] A. Raj, J.A. Kumar, P. Bansal, A multicriteria decision making approach to study barriers to the adoption of autonomous vehicles, *Transportation research part A: policy and practice* 133 (2020) 122–137.
- [8] E. Namazi, R.N. Holthe-Berg, C.S. Lofsberg, J. Li, Using vehicle-mounted camera to collect information for managing mixed traffic, in: 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), IEEE, 2019, pp. 222–230.
- [9] E. Namazi, J. Li, R. Mester, C. Lu, Identifying and counting vehicles in multiple lanes by using a low-cost vehicle-mounted sensor for intelligent traffic management systems, in: *International Conference on Hybrid Artificial Intelligence Systems*, Springer, 2020, pp. 598–611.
- [10] Movable-Type, Calculate distance, bearing and more between latitude/longitude points. <https://www.movable-type.co.uk/scripts/latlong.html>.
- [11] E. Javanmardi, Y. Gu, M. Javanmardi, S. Kamijo, Autonomous vehicle self-localization based on abstract map and multi-channel lidar in urban area, *IATSS research* 43 (1) (2019) 1–13.
- [12] A. Chehri, N. Quadar, R. Saadane, Survey on localization methods for autonomous vehicles in smart cities, in: *Proceedings of the 4th International Conference on Smart City Applications*, 2019, pp. 1–6.
- [13] N. Acosta, J. Toloza, Techniques to improve the gps precision, *International Journal of Advanced Computer Science and Applications* 3 (8).
- [14] Z. Huang, S. Qiao, N. Han, C.-A. Yuan, X. Song, Y. Xiao, Survey on vehicle map matching techniques, *CAAI Transactions on Intelligence Technology* 6 (1) (2021) 55–71.
- [15] M.S. Iftekhar, N. Saha, Y.M. Jang, Stereo-vision-based cooperative-vehicle positioning using occ and neural networks, *Optics Communications* 352 (2015) 166–180.
- [16] J. Hayakawa, B. Dariush, Ego-motion and surrounding vehicle state estimation using a monocular camera, in: *IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2019, pp. 2550–2556.
- [17] J. Lee, Intervehicle distance estimation through camera images, *Journal of Electronic Imaging* 27 (6) (2018) 063001.
- [18] D.-Y. Huang, C.-H. Chen, T.-Y. Chen, W.-C. Hu, K.-W. Feng, Vehicle detection and inter-vehicle distance estimation using single-lens video camera on urban/suburb roads, *Journal of Visual Communication and Image Representation* 46 (2017) 250–259.
- [19] J.L. Giesbrecht, H.K. Goi, T.D. Barfoot, B.A. Francis, A vision-based robotic follower vehicle, in: *Unmanned Systems Technology XI*, Vol. 7332, International Society for Optics and Photonics, 2009, p. 733210.
- [20] E. Namazi, R. Mester, C. Lu, M.M. Log, J. Li, Improving vehicle localization with two low-cost gps receivers, in: *The sixth smart city applications international conference*, 2021.
- [21] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.
- [22] J. Redmon, A. Farhadi, Yolov3: An incremental improvement, *arXiv preprint arXiv:1804.02767*.
- [23] Yolo: Real-time object detection. <https://pjreddie.com/darknet/yolo/>.
- [24] D. Wang, C. Li, S. Wen, Q.-L. Han, S. Nepal, X. Zhang, Y. Xiang, Daedalus: Breaking nonmaximum suppression in object detection via adversarial examples, *IEEE Transactions on Cybernetics*.
- [25] J. Redmon, Pytorch-yolov3-kitti. <https://github.com/packyan/PyTorch-YOLOv3-kitti>.
- [26] Gopro, Hero7 field of view (fov) information. [https://gopro.com/help/articles/question\\_answer/hero7-field-of-view-fov-information?sf96748270=1..](https://gopro.com/help/articles/question_answer/hero7-field-of-view-fov-information?sf96748270=1..)

- [27] L. Ding, A. Goshtasby, On the canny edge detector, *Pattern Recognition* 34 (3) (2001) 721–725.
- [28] C. Galamhos, J. Matas, J. Kittler, Progressive probabilistic hough transform for line detection, in: Proceedings. 1999 IEEE computer society conference on computer vision and pattern recognition (Cat. No PR00149), Vol. 1, IEEE, 1999, pp. 554–560..
- [29] J. Matas, C. Galambos, J. Kittler, Robust detection of lines using the progressive probabilistic hough transform, *Computer vision and image understanding* 78 (1) (2000) 119–137.
- [30] Thales theorem and figure. <https://sites.math.washington.edu/king/coursedir/m444a02/class/10-21-thales.html>..
- [31] Euclidean distance. [https://en.wikipedia.org/wiki/Euclidean\\_distance](https://en.wikipedia.org/wiki/Euclidean_distance)..
- [32] Trigonometry. <https://en.wikipedia.org/wiki/Trigonometry>..



**Chaoru Lu** received the Ph.D. degree in civil engineering from Iowa State University, in 2017. He is currently an Associate Professor with the Department of Civil Engineering and Energy Technology, Oslo Metropolitan University, Oslo, Norway. His research interests include connected and automated vehicles, transport electrification, and intelligent transportation systems.



**Elnaz Namazi** received the Ph.D. degree in computer science from the Norwegian University of Science and Technology (NTNU) in 2022. Her research interests include computer vision, digitalization, intelligent transportation systems, and connected and automated vehicles.



**Jingyue Li** received the PhD degree in software engineering from the Department of Computer Science, Norwegian University of Science and Technology (NTNU), in 2006. He is a Professor in the Computer Science Department, NTNU. His recent research interests include software engineering, software security, blockchain, and big data.



**Rudolf Mester** received the Dr.-Ing. degree from RWTH Aachen in 1989. He was a Professor in Applied Physics and later in Computer Science with Goethe University Frankfurt from 1995. He is a Professor in computer science and a member of the Norwegian Open AI Laboratory since October 2018. He is an expert on environment sensing for autonomous systems, and has also experience from 7 years in industrial advanced development and research (Bosch, Philips). His fields of work comprise signal processing, state estimation, computer vision, and statistical data analysis. Since about 5 years his research focuses on machine learning, dependable AI, and autonomous systems.