IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

Online Camera Orientation Calibration Aided by A High-Speed Ground-View Camera

Junzhe Su[1], Masahiro Hirano[1], Yuji Yamakawa[1]

***Abstract*—This paper proposes an online method for calibrating the orientation of cameras mounted on vehicles. To calibrate the orientation of the target camera relative to the vehicle, we propose using a high-speed vision sensor, focused on the ground, in conjunction with the target camera. First, the high-speed camera's planar motion parallel to ground plane and target camera's motion are estimated by a semi-dense approach and visual odometry method, respectively. Then, the motions are used to calibrate the target camera's orientation through nonlinear optimization based on the invariance constraint of the extrinsic parameters and the nonholonomic constraint of the vehicle. Unlike traditional methods, this approach does not depend on artificial features such as lane markings and utilizes ground information more efficiently, making it applicable in broader scenarios. Simulation and field tests have demonstrated that the target camera orientation calibration errors are approximately 1°, even on a bumpy road, affirming the accuracy and robustness of the proposed method.**

***Index Terms*—Calibration and identification, wheeled robots, high-speed vision, onboard vehicular camera, nonholonomic constraint.**

I. Introduction

Since the advent of automobiles, efforts have been made to enhance their safety and comfort. From advanced driver-assistance systems (ADASs) to autonomous driving technology, vehicles are becoming increasingly intelligent and efficient. A variety of sensors, including cameras, are essential to perceive the surrounding environment for vehicles' safe and proper operation. Cameras play a crucial role in detecting objects like vehicles, traffic signs, lane markings, and vulnerable road users in the vicinity of the ego vehicle. In autonomous emergency braking (AEB) systems, cameras identify potential collision objects and calculate their distance.

The accurate and robust perception of the environment by cameras requires correct determination of the camera's pose relative to the ego vehicle coordinate system, known as the extrinsic parameter. Inaccurate extrinsic parameters can lead to incorrect information processing in subsequent modules, such as route planning and decision making (refer to Section II for details). Over time, factors like mechanical vibrations, temperature variations, payload variety, vehicle repairs, and part replacements may alter a camera's extrinsic parameters, necessitating periodic recalibration to maintain accuracy.

Calibration of camera extrinsic parameters can be achieved through offline or online approaches. Offline methods depend on pre-defined patterns such as checkerboards [1] and may need specialized facilities, making them inaccessible for typical drivers. Online methods, on the other hand, utilize information from the vehicle's surroundings, captured by cameras during operation, to estimate extrinsic parameters automatically. This kind of approach is more convenient, requiring less human intervention, and can also simplify sensor mounting during vehicle production [2].

Online camera calibration methods have been a prevalent area of study. Most research has focused on using artificial features to calibrate cameras. In one approach, detailed in [3], cameras were calibrated by assuming that part of the ego vehicle remains within the camera frame, thus enabling the use of photographed vehicle information. Studies in [4-7] employed lane markings to calibrate the camera's extrinsic parameters, while vanishing points (VPs) derived from environmental line structures were used in [8-10]. The accuracy of these methods, however, largely relies on the quality of the artificial features. If the cameras cannot observe the ego vehicle or if the environment lacks sufficient lane markings and line structures, the calibration accuracy significantly decreases. Moreover, for methods using VPs to calibrate cameras, the line segments detected in images could contain considerable noise data, which also affects these methods' overall effectiveness.

Some studies have shifted focus towards online camera calibration using natural features extractable from ordinary scenes. Research in [11, 12] extracted feature points on the ground plane to estimate camera extrinsic parameters. These studies face a critical limitation: the ground plane consists of mostly high frequency, self-similar, speckly texture, which results in difficulty in reliable feature matching or even extraction on it, and the accuracy of these methods cannot be guaranteed.

Fusing information from other types of sensors is helpful in solving existing problems in the aforementioned research. Studies in [13, 14] applied synchronized vehicle poses, estimated by an inertial measurement unit (IMU), to support the calibration of camera extrinsic parameters. In [15], wheel odometry measurements were integrated with visual odometry

Manuscript received: March 20, 2023; Revised June 7, 2023; Accepted July 26, 2023.

This paper was recommended for publication by Editor Lucia Pallottino upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by China Scholarship Council (Grant number: 202106260030).

[1]Junzhe Su, Masahiro Hirano and Yuji Yamakawa are with Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan sujunzhe@iis.u-tokyo.ac.jp

Digital Object Identifier (DOI): see top of this page.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

measurements to form a hypergraph, and graph optimization was conducted to estimate the camera's extrinsic parameters. However, these methods still possess inherent limitations. Synchronization between different types of sensors introduces complexity, and the drift in IMU and wheel odometry measurements, coupled with the effect of slip between the tire and ground on wheel odometry, negatively impact calibration accuracy of these methods.

This study proposes a novel online camera extrinsic parameters calibration method. Specially, our method focuses on online calibration of camera orientation relative to vehicle. Considering the vehicle's movement as planar motion on the ground plane, the spatial relationship between the camera and the ground plane becomes vital for calibration. However, as noted earlier, the noisy texture of road surface often hinders sufficient information extraction. To address this, the proposed method employs a high-speed ground-view camera in conjunction with a target camera. We utilize a high-speed camera based on the observation that its high frame rate allows for capturing images with small displacement between them. Consequently, a semi-dense approach proposed in a previous work [16], which is capable of analyzing noisy or even poorly textured road surfaces, can be used to estimate the ground view camera's planar motion (i.e., the motion of its coordinate system projection on the road). Simultaneously, the target camera motion can be estimated by the visual odometry method. Subsequently, these motions are included into a nonlinear optimization process based on the invariance constraint of the extrinsic parameters and the vehicle's nonholonomic constraint to estimate the target camera's orientation relative to the vehicle. Simulation and field tests including comparative evaluation experiments were conducted to verify the validity of the proposed algorithm.

The core contribution of this study is the development of a novel online camera calibration process utilizing camera motions which are estimated by efficiently extracting information from ground texture and natural features in the surrounding environment of ego vehicle. Thus, it obviates the need for manmade structures and expands its application range. The ability of proposed method to handle various textures was proven by the simulation test. Compared with conventional methods, owing to its inherent characteristics, the proposed method has fewer constraints on the installation position of the onboard camera to be calibrated, it can be installed on the front windshield, the rear, or other locations provided its field of view contains adequate information for motion estimation. This method is also adaptable for calibrating other types of sensors like LiDAR, which can estimate their own motions.

We believe the integration of high-speed camera is both reasonable and meaningful, as it can be used not only for camera calibration but also be utilized to detect lane markings or function as visual odometry. Nowadays, cost of implementing high-speed cameras continues to decrease, aligning with trends observed in other sensors.

II. Sensitivity Analysis

This section presents a sensitivity analysis of the vehicle's environmental perception function with respect to the camera extrinsic parameters.

When objects are detected in the camera coordinate system, the extrinsic parameters including translation and rotation are used to transform them into the vehicle coordinate system for subsequent processing. When errors arise in the translation, regardless of the distance between the obstacles and ego vehicle, the error in the detected obstacle position is the same as the camera translation error. Conversely, when errors occur in the camera rotation, it introduces a diminishing accuracy in obstacle localization as the distance to the obstacles increases, which can be intuitively understood through the arc length formula $l = \alpha r$, where l, α, and r represent the arc length, central angle, and radius, respectively. Therefore, the proposed method focuses on calibrating the camera orientation.

As depicted in Fig. 1, three obstacle vehicles (Targets 1-3) were positioned in front of the ego vehicle in the simulation environment. A camera, affixed to the front windshield, gauged the positions of these obstacles relative to the ego vehicle. The ego vehicle's coordinate system was right-handed, with the x-axis oriented forward and the y-axis to the left. With a 1° error in the camera rotation angles around all three axes, the coordinate errors for Targets 1–3 are confined within 0.55 m. An increase to a 2° error in the camera rotation angles causes the errors in the y- and z-coordinates of Target 3 to grow to approximately 1 m. A further increase to a 3° error results in the y- and z-coordinate errors for Target 2 increase to approximately 1 m, and those of Target 3 to approximately 1.5 m. These results indicate that errors in the camera rotation angles could lead to overlooking a front obstacle in the ego vehicle's lane, potentially causing grave traffic accidents. Using

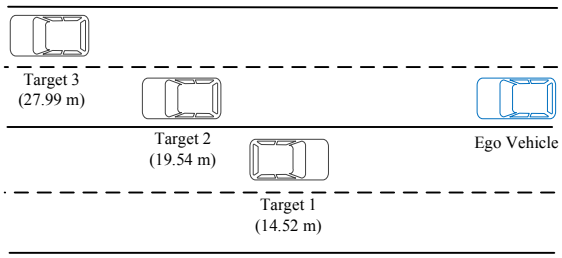

Fig. 1. Distribution of obstacles and ego vehicle in simulation environment. The numbers below each target represent their distance relative to the ego vehicle.

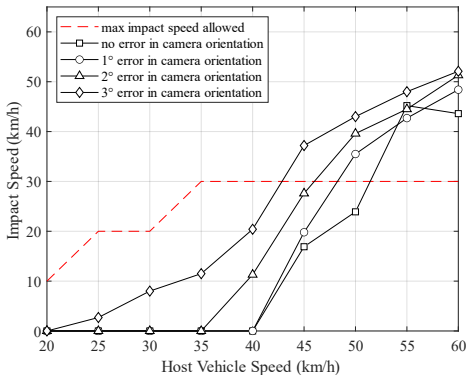

Fig. 2. Results of AEB system assessment. The horizontal axis represents the host vehicle test speed, and the vertical axis represents the speed at which the host vehicle collides with obstacles. The red dotted line is the max impact speed allowed for each test speed.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

the same simulation environment, a more specific experiment was conducted with a car equipped with an AEB system using data from a camera and a radar. The AEB functionality was assessed according to the Euro NCAP assessment protocol [17]. As illustrated in Fig. 2, although the impact speed of host vehicle at 55 km/h with no error in camera orientation is higher than that with 1° and 2° error due to modeling errors of AEB system in simulation software, as the camera orientation error increases, the AEB system performance deteriorates, as evidenced by a rise in the host vehicle's impact speed at nearly every test speed. We also added 0.05 m error to camera translation relative to vehicle and assessed the influence of it on AEB functionality. The results show that with 0.05 m error in camera translation, the impact speed of host vehicle does not have much difference from that without errors (the maximum difference is 0.7 km/h), which shows that it is reasonable to neglect the translation error.

Based on the above quantitative analysis, calibrating the orientation of a vehicle-mounted camera is necessary to ensure camera-based systems on the vehicle function properly; additionally, the calibration error should be limited to 1° for optimal results.

III. Geometry and Camera Orientation Calibration Method

The ground plane in the vicinity of the ego vehicle is an essential information source. Because the vehicle motion can be considered as planar motion on the ground plane, the spatial relationship between the camera and ground plane can be used to calibrate the camera orientation relative to the vehicle. This study introduces a ground view camera with a high frame rate (like 200 Hz) as an intermediary to assist with target camera orientation calibration. First, utilizing the high-speed characteristics of the ground-view camera, we employ a semi-dense approach to analyze road images captured by the camera without extracting feature points. This enables the camera orientation relative to the ground plane and its planar motion on the road surface to be estimated simultaneously. Concurrently, we estimate the target camera's motion using the visual odometry method. Following this, we estimate the target camera's orientation relative to the projection of the ground-view camera on the road. This is achieved using the invariant constraint of the extrinsic parameters combined with the target camera motion and ground-view camera planar motion. We can then ascertain the yaw angle of the ground-view camera relative to the vehicle coordinate system by using the nonholonomic constraint of the vehicle. Ultimately, the target camera orientation relative to the vehicle, including the roll, pitch, and yaw angles, can be obtained from the results of these preceding steps.

A. Configuration and Geometry

In this paper, the target camera to be calibrated is denoted as camera$_a$, where 'a' indicates 'aim'; the intermediary camera facing the ground is denoted as camera$_i$. As shown in Fig. 3, there are four coordinate systems: $\mathbf{C}_i$ is for camera$_i$, $\mathbf{C}_a$ is for camera$_a$, $\mathbf{C}_g$ is the projection of $\mathbf{C}_i$ on the ground (see the following section for further details), and $\mathbf{C}_v$ is for the vehicle. In the vehicle coordinate system $\mathbf{C}_v$, the z-axis points in the forward direction, the x-axis is perpendicular to it on a plane parallel to the ground, and the origin lies in the center of the back-wheel axis on the ground. The transformations between coordinate systems are defined as shown in Fig. 3, and $\mathbf{T}_{va}$ represents camera$_a$'s extrinsic parameters relative to the vehicle. $\mathbf{T}_{va}$ is composed of rotation $\mathbf{R}_{va}$ and translation $\mathbf{t}_{va}$. Because our target is to calibrate the camera orientation, only $\mathbf{R}_{va}$ is estimated in the proposed method, and $\mathbf{t}_{va}$ remains unchanged from the initial value. The framework of the proposed method is shown in Fig. 4: first, the motions of the two cameras are estimated in Step 1; next, $\mathbf{R}_{ga}$ is estimated using the invariance constraint of the extrinsic parameters in Step 2; then, $\mathbf{R}_{vg}$ is estimated using the nonholonomic constraint of the vehicle in Step 3; finally, $\mathbf{T}_{va}$ can be calculated using (1), with $\mathbf{t}_{ga}$ and $\mathbf{t}_{vg}$ remaining unchanged from the initial values, and camera$_a$ orientation $\mathbf{R}_{va}$ can be extracted from $\mathbf{T}_{va}$.

$$\mathbf{T}_{va} = \mathbf{T}_{vg}\mathbf{T}_{ga} \tag{1}$$

B. Step 1: Estimation of Camera Motions

Because camera$_i$ acts as an intermediary to help calibrate camera$_a$ orientation $\mathbf{R}_{va}$, it is necessary first to estimate camera$_i$

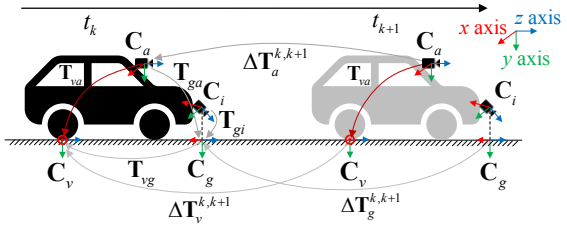

Fig. 3. Geometry of the problem with temporal transformations from time k to time k+1 and extrinsic transformations. $\mathbf{T}_{ga}$, $\mathbf{T}_{gi}$, $\mathbf{T}_{vg}$ and $\mathbf{T}_{va}$ remain unchanged so they are not all labeled at time k+1.

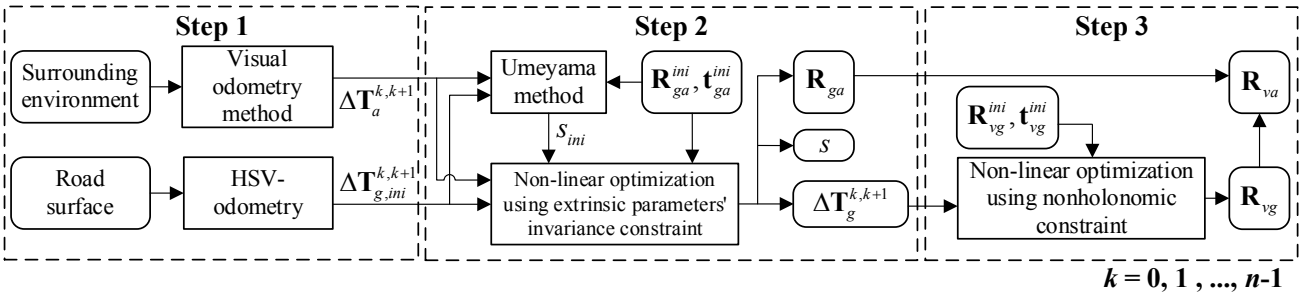

Fig. 4. Framework of proposed method. k = 0, 1, ..., n-1 means there are n+1 frames in total, from which n relative motions between every two consecutive frames are calculated and input to the algorithm.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

planar motion $\Delta\mathbf{T}_g^{k,\,k+1}$ and camera$_a$ motion $\Delta\mathbf{T}_a^{k,\,k+1}$ between time k and time k+1. These can then be used to estimate camera$_a$ orientation $\mathbf{R}_{va}$ based on the geometry transformations shown in Fig. 3. Camera$_a$ motion $\Delta\mathbf{T}_a^{k,\,k+1}$ can be estimated by using the visual odometry method with images captured by camera$_a$.

To estimate camera$_i$ planar motion $\Delta\mathbf{T}_g^{k,\,k+1}$, an algorithm called high-speed vision odometry (HSV-odometry) proposed in a previous study is used. As shown in Fig. 3, $\mathbf{C}_g$ is the projection of $\mathbf{C}_i$ on the ground; its y-axis is perpendicular to the ground plane, and its origin is on the ground. Because ground-view camera$_i$ has a high frame rate, the planar motion is denoted as $\Delta\mathbf{T}_g^{j,\,j+1}$, and the time interval between time j and time j+1 corresponds to the camera$_i$ frame rate. $\mathbf{T}_{gi}$ and $\Delta\mathbf{T}_g^{j,\,j+1}$ are represented by (2) and (3), respectively, where the first three elements represent translation in the order of x, y, and z, and the last three elements represent the pitch angle around the x-axis, yaw angle around the y-axis, and roll angle around the z-axis, respectively. In particular, h in $\mathbf{T}_{gi}$ is the height of $\mathbf{C}_i$ from the ground and is given as prior information in the algorithm. With a given value of h, the real scale of the estimated $x_g^{j,\,j+1}$ and $z_g^{j,\,j+1}$ can be obtained.

$$\mathbf{T}_{gi} = \left(0,\, 0,\, -h,\, pitch_{gi},\, 0,\, roll_{gi}\right) \tag{2}$$

$$\Delta\mathbf{T}_g^{j,\,j+1} = \left(x_g^{j,\,j+1},\, 0,\, z_g^{j,\,j+1},\, 0,\, yaw_g^{j,\,j+1},\, 0\right) \tag{3}$$

Images of the ground surface at time j and time j+1 captured by camera$_i$ are marked as img_j and img_{j+1}, respectively, and are input to the algorithm. By adopting a semi-dense approach, a set of patches is chosen on img_j and img_{j+1} to compute their apparent motion via an image-registration method called phase-only correlation [18]. Subsequently, optimization is performed by employing a robust geometric bundle adjustment to obtain $pitch_{gi}$ and $roll_{gi}$ in $\mathbf{T}_{gi}$ and $\Delta\mathbf{T}_g^{j,\,j+1}$ simultaneously. The small translation between img_j and img_{j+1} under the high frame rate of camera$_i$ guarantees the feasibility of the algorithm. This image registration method enables the algorithm to be applied even on poor texture ground. Finally, $\Delta\mathbf{T}_g^{k,\,k+1}$ is calculated using a sequence of incremental motions composed of $\Delta\mathbf{T}_g^{j,\,j+1}$. Further details can be found in [16].

C. Step 2: Estimation of Target Camera Orientation Relative to Ground

When a vehicle travels on a regular flat road without maneuvering motions, such as acceleration or turning at high speed, the extrinsic parameters of the onboard cameras can be approximated as invariant. According to the geometric relationship, we have

$$\mathbf{T}_{ga}\Delta\mathbf{T}_a^{k,\,k+1} = \Delta\mathbf{T}_g^{k,\,k+1}\mathbf{T}_{ga}, \tag{4}$$

where $\Delta\mathbf{T}_a^{k,\,k+1}=[\mathbf{R},\,\mathbf{t}]$ ($\mathbf{R}\in\mathrm{SO}(3)$, $\mathbf{t}\in\mathbb{R}^3$) is the motion of $\mathbf{C}_a$ and $\Delta\mathbf{T}_g^{k,\,k+1}=[\mathbf{R},\,\mathbf{t}]$ ($\mathbf{R}\in\mathrm{SO}(2)$, $\mathbf{t}\in\mathbb{R}^2$) is the planar motion of $\mathbf{C}_g$ between time k and time k+1.

Because monocular cameras cannot provide actual scale information, $\mathbf{T}_{ga}$ should be expressed by the Lie group of similarity transforms SIM(3), i.e., Euclidean motions with uniform scaling [19]. Let us consider SIM(3) in the following form.

$$\mathrm{SIM}(3)=\left\{\mathbf{S}=\begin{bmatrix} s\mathbf{R} & \mathbf{t} \\ 0_{1\times3} & 1 \end{bmatrix}\in\mathbb{R}^{4\times4}\,|\,s\in\mathbb{R}^{+},\mathbf{R}\in\mathrm{SO}(3),\mathbf{t}\in\mathbb{R}^{3}\right\} \tag{5}$$

$\mathbf{T}_{ga}$ can be then expressed by

$$\mathbf{T}_{ga}=\begin{bmatrix} s\mathbf{R}_{ga} & \mathbf{t}_{ga} \\ 0_{1\times3} & 1 \end{bmatrix}, \tag{6}$$

where the product of the rotation matrix and scale $s\mathbf{R}_{ga}$ is further denoted as $\mathbf{R}_{ga}^{s}$.

Although $\Delta\mathbf{T}_g^{k,\,k+1}$ is in SE(2), it can be referred to in three dimensions by raising it into SE(3):

$$\Delta\mathbf{T}_g^{k,\,k+1}=\begin{bmatrix} \cos\theta_g^{k,\,k+1} & 0 & \sin\theta_g^{k,\,k+1} & x_g^{k,\,k+1} \\ 0 & 1 & 0 & 0 \\ -\sin\theta_g^{k,\,k+1} & 0 & \cos\theta_g^{k,\,k+1} & z_g^{k,\,k+1} \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{7}$$

where $\theta_g^{k,\,k+1}$ represents the yaw angle around the y-axis.

Then, $\Delta\mathbf{T}_a^{k,\,k+1}$ and $\Delta\mathbf{T}_g^{k,\,k+1}$ can be regarded as members of SIM(3) with scale s = 1. Because there are errors in the initial value of $\mathbf{T}_{ga}$ and the estimated $\Delta\mathbf{T}_g^{k,\,k+1}$ and $\Delta\mathbf{T}_a^{k,\,k+1}$, (4) does not hold exactly. Consequently, we define the error $\mathbf{E}_{k,k+1}$ as

$$\mathbf{E}_{k,k+1}=\Delta\mathbf{T}_a^{k,\,k+1}{}^{-1}\mathbf{T}_{ga}{}^{-1}\Delta\mathbf{T}_g^{k,\,k+1}\mathbf{T}_{ga}. \tag{8}$$

Because $\mathbf{E}_{k,k+1}$ is a member of SIM(3), its logarithm $\boldsymbol{e}_{k,k+1}$ belongs to the Lie algebra $\mathfrak{sim}(3)$:

$$\boldsymbol{e}_{k,k+1}=\log\left(\Delta\mathbf{T}_a^{k,\,k+1}{}^{-1}\mathbf{T}_{ga}{}^{-1}\Delta\mathbf{T}_g^{k,\,k+1}\mathbf{T}_{ga}\right). \tag{9}$$

Then cost function f can be written as

$$f=\sum_{k=0}^{n-1}\boldsymbol{e}_{k,k+1}^{\mathrm{T}}\Omega_{k,k+1}\boldsymbol{e}_{k,k+1}, \tag{10}$$

where $\Omega_{k,k+1}$ is the information matrix of measurement $\Delta\mathbf{T}_a^{k,\,k+1}$. $\Delta\mathbf{T}_a^{k,\,k+1}$ is chosen as the measurement because $\Delta\mathbf{T}_g^{k,\,k+1}$ is estimated by observing a small local area; thus, it is less robust than $\Delta\mathbf{T}_a^{k,\,k+1}$, which is obtained from images captured by camera$_a$ with a broader field of view. Our goal is to find the set of parameters $\left(\mathbf{R}_{ga}^{s},\,\Delta\mathbf{T}_g^{k,\,k+1}\right)$ that minimizes the cost function f. We employ the factor graph shown in Fig. 5 to solve the optimization problem. In this graph, the circle nodes represent the variables to be optimized, and the square factors represent the measurements obtained from visual odometry using images from camera$_a$. Each time new estimation of $\Delta\mathbf{T}_a^{k,\,k+1}$ and $\Delta\mathbf{T}_g^{k,\,k+1}$ are available, an $\mathbf{R}_{ga}^{s}$ node and a new $\Delta\mathbf{T}_g^{k,\,k+1}$ node will be connected with a new edge containing factor $\Delta\mathbf{T}_a^{k,\,k+1}$. We used g2o [20] for the optimization to obtain s, $\mathbf{R}_{ga}$ and $\Delta\mathbf{T}_g^{k,\,k+1}$.

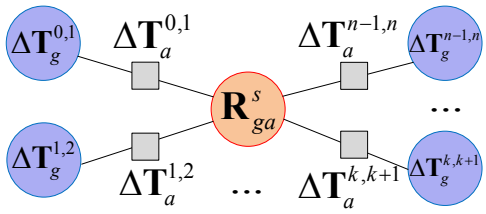

Fig. 5. Factor graph of $\mathbf{R}_{ga}$ estimation.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

Because the initial extrinsic parameters of camera$_i$ and camera$_a$ relative to the vehicle are given as prior information, the initial values of $\mathbf{R}_{ga}$ and $\mathbf{t}_{ga}$ can be calculated from them. Subsequently, by transforming the initial value of $\Delta\mathbf{T}_g^{k,\,k+1}$ into $\mathbf{C}_a$ using the initial values of $\mathbf{R}_{ga}$ and $\mathbf{t}_{ga}$, the initial value of scale s can be estimated by combining $\Delta\mathbf{T}_a^{k,\,k+1}$ using the Umeyama method proposed in [21].

D. Step 3: Estimation of Ground View Camera Yaw Angle Relative to Vehicle

A vehicle is a nonholonomic system, and thus, is subject to motion constraints. Inspired by the method in [22], the nonholonomic constraint is used to estimate the relative orientation of $\mathbf{C}_g$ to $\mathbf{C}_v$, which is denoted as $\mathbf{R}_{vg}$. Unlike in [22], where the absolute trajectory of the camera was used, this study uses relative motion between every two consecutive frames to reduce the influence of accumulated errors and guarantee the estimation accuracy of $\mathbf{R}_{vg}$.

From Fig. 3, the incremental motion of vehicle $\Delta\mathbf{T}_v^{k,\,k+1}$ can be calculated using (11).

$$\Delta\mathbf{T}_v^{k,\,k+1}=\mathbf{T}_{vg}\Delta\mathbf{T}_g^{k,\,k+1}\mathbf{T}_{vg}^{-1} \tag{11}$$

In this study, vehicle motion is restricted to the ground plane; thus, $\Delta\mathbf{T}_v^{k,\,k+1}$ and $\Delta\mathbf{T}_g^{k,\,k+1}$ both represent 2D motion. Without loss of generality, $\mathbf{C}_v$ and $\mathbf{C}_g$ are rotated such that their x-axes point to the front and y-axes point to the left compared to their original orientation. $\mathbf{R}_{vg}$ can then be regarded as a rotation with θ_{vg} degrees around the z-axis, which is perpendicular to the ground plane. Let $\Delta\mathbf{T}_g^{k,\,k+1}=\left(x_g^{k,k+1},y_g^{k,k+1},\theta_g^{k,k+1}\right)$, $\mathbf{T}_{vg}=\left(x_{vg},y_{vg},\theta_{vg}\right)$, then the nonholonomic constraint can be expressed as

$$\dot{x}_v^{k,k+1}\sin\theta_v^{k,k+1}-\dot{y}_v^{k,k+1}\cos\theta_v^{k,k+1}=0, \tag{12}$$

where $\Delta\mathbf{T}_v^{k,\,k+1}=\left(x_v^{k,k+1},y_v^{k,k+1},\theta_v^{k,k+1}\right)$ and the vehicle velocities between time k and time k+1 are $\dot{x}_v^{k,k+1}$ and $\dot{y}_v^{k,k+1}$.

Because cameras are rigidly linked to the vehicle when the vehicle is moving and $\Delta\mathbf{T}_g^{k,\,k+1}$ and $\Delta\mathbf{T}_v^{k,\,k+1}$ are defined in $\mathbf{C}_g$ and $\mathbf{C}_v$, respectively, the following formula can be obtained:

$$\theta_v^{k,k+1}=\theta_g^{k,k+1}. \tag{13}$$

Since there are errors in $\Delta\mathbf{T}_g^{k,\,k+1}$ and $\mathbf{T}_{vg}$, (12) cannot hold exactly. By combining (11), (12), and (13), the error based on nonholonomic constraint $\boldsymbol{e}_{k,k+1}(\theta_{vg})$ is calculated as follows:

$$\begin{aligned}\boldsymbol{e}_{k,k+1}&=\dot{x}_v^{k,k+1}\sin\theta_v^{k,k+1}-\dot{y}_v^{k,k+1}\cos\theta_v^{k,k+1}\\&=\dot{y}_g^{k,k+1}\cos\left(\theta_{vg}-\theta_g^{k,k+1}\right)+\dot{x}_g^{k,k+1}\sin\left(\theta_{vg}-\theta_g^{k,k+1}\right)+\dot{\theta}_g^{k,k+1}x_{vg}.\end{aligned} \tag{14}$$

As the motion between two consecutive frames is small, we can assume $\dot{x}_g^{k,k+1}$, $\dot{y}_g^{k,k+1}$ and $\dot{\theta}_g^{k,k+1}$ are constant during motion. The initial values of x_{vg} and θ_{vg} are calculated from the initial extrinsic parameters of camera$_i$. Then, θ_{vg} can be estimated by minimizing the cost function, as shown in (15). As in Step 2, g2o [20] is used to solve this least-squares problem.

$$\theta_{vg}=\underset{\theta_{vg}}{\operatorname{argmin}}\sum_{k=0}^{n-1}\boldsymbol{e}_{k,k+1}^2 \tag{15}$$

Once $\mathbf{R}_{ga}$ and $\mathbf{R}_{vg}$ have been estimated, the orientation of camera$_a$ relative to the vehicle $\mathbf{R}_{va}$ can be extracted from $\mathbf{T}_{va}$ using (1). As shown in [23], camera$_a$ orientation cannot be fully calibrated when the vehicle only experiences a straight line or circular motion. In practice, these degenerate cases can be easily avoided by varying the vehicle velocity and trajectory geometry.

IV. Experiment Results and Analysis

The experiment consisted of both simulation and field tests.[1] A comparison was conducted between proposed method and a state-of-the-art online camera orientation calibration method SensorX2car [24], which is based on VPs detection. For the simulation tests, CARLA [25] was used to construct the simulation environment, and tests were performed using the camera configuration shown in Fig. 6. In the field test, a mobile robot 1/10 the size of an average car was used, and it traveled on a regular road. To evaluate the validity of the proposed method, errors were added to the ground truth of the extrinsic parameters of the cameras, and they were used as initial values for the calibration algorithm. The error for the camera rotation was 5° and that for the translation was 0.03 m. The errors were added to the rotation angles around the three axes and the translation in the three coordinates. Considering the accuracy and robustness of the algorithm, ORB-SLAM2 [26] was selected to estimate the trajectory of the camera to be calibrated.

A. Simulation Test

In the simulation environment, a sedan equipped with two cameras traveled on a normal road. The frame rate of camera$_a$ was 25 Hz (resolution: 1280×1024 pixels). The height of camera$_i$ from the ground was 0.6 m, and its frame rate was 250 Hz (resolution: 800×600 pixels). To compare proposed method with SensorX2car method, two image sequences lasting 13–15 s, which were denoted as curve sequence and mixed sequence, were collected. In the curve sequence, ego car travelled on a curve road without straight segments; in the mixed trajectory, ego car travelled on a straight road and performed several lane

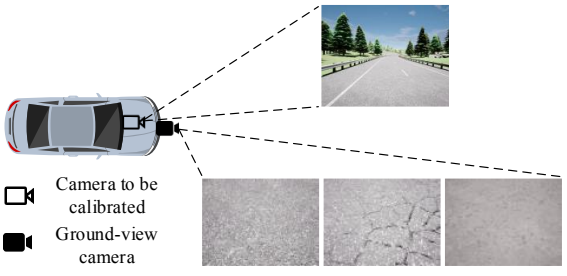

Fig. 6. Camera configuration in simulation test and sample images captured by cameras. In the ground-view camera's images, from left to right are normal asphalt road, ruined asphalt road and concrete road, respectively. RGB images are converted to grayscale images for the algorithm.

[1]Experiment data are available on http://www.hfr.iis.u-tokyo.ac.jp/research/OnlineCalibration/index-e.html

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

TABLE I. SIMULATION TEST RESULTS OF TWO DIFFERENT METHODS[a]

		SensorX2car [24]			Proposed method		
		Yaw	*Pitch*	*Roll*	*Yaw*	*Pitch*	*Roll*
Curve trajectory	Result (Error)	2.512 (3.512)	-1.397 (-0.397)	0.148 (2.148)	-0.932 (0.068)	-0.998 (0.002)	-1.307 (0.693)
Mixed trajectory	Result (Error)	1.491 (2.491)	-1.302 (-0.302)	-1.086 (0.914)	-0.912 (0.088)	-0.997 (0.003)	-1.935 (0.065)

[a]The unit of all results is degree. Yaw angle, pitch angle and roll angle correspond to *y*-axis, *x*-axis and *z*-axis, respectively. The rules also apply to following other tables.

changes, therefore, the trajectory was composed of straight segments and curve segments. For proposed method, to eliminate the influence of the difference between each image sequence duration, every image sequence was subdivided into multiple overlapping sub-sequences with the same number of frames, starting with every new frame. Calibration was conducted on each subsequence, and the final calibration result for each image sequence was the average of each subsequence result. For SensorX2car method, because it used only single image to calibrate camera orientation, the final calibration result for each image sequence was the average of single image result. The calibrated extrinsic parameters were then compared with the ground truth obtained from the simulation software.

The simulation test results of two methods are listed in Table I, where all three angles of $\mathbf{R}_{va}$ are shown. As shown in the table, proposed method outperforms SensorX2car method in both two trajectories with smaller calibration errors of less than 0.7°. Due to the limitation of the assumption that while the vehicle is driving straightly, the direction of the vehicle can be regarded as the direction of the road, although SensorX2car method can extract useful driving segments, it has poor performance in curve trajectory compared with that in mixed trajectory with straight segments, which is shown by calibration errors of pitch angle and roll angle less than 1° and around 1° lower yaw angle calibration error in mixed trajectory. Based on the above analysis, we can see that the proposed method is more accurate and robust under different kinds of trajectories compared with SensorX2car method. Moreover, although there are errors in the translation, the test results show that the calibration accuracy is not significantly affected by the errors, which also verifies the robustness of the proposed method.

Because the analysis of road surface is an essential part of proposed method, the impact of different road surface textures was evaluated. Ego car was traveling on the same curve road as previous test while images were captured. Three different kinds of road surfaces including normal asphalt road, ruined asphalt road and concrete road shown in Fig. 6 were used to form three image sequences. Table II lists the simulation test results. From the results we can see, different road surface textures do not have much impact on the calibration accuracy of proposed method with all calibration errors within 0.7°, which shows the robustness and sufficient accuracy of proposed method on different road surface textures.

TABLE II. SIMULATION TEST RESULTS OF DIFFERENT ROAD SURFACE TEXTURES

		Proposed method		
		Yaw	*Pitch*	*Roll*
Normal asphalt road	Result	-0.932	-0.998	-1.307
	Error	0.068	0.002	0.693
Ruined asphalt road	Result	-0.964	-1.001	-1.477
	Error	0.036	-0.001	0.523
Concrete road	Result	-0.893	-1.004	-1.645
	Error	0.107	-0.004	0.355

We also test the performance of proposed method on different vehicle speeds with current camera configuration. Ego car was traveling on the same curve road as previous test while images were captured and vehicle speeds were 8 m/s and 12 m/s, respectively. The calibration errors on two speeds do not change much and are all within 0.7°, which shows that the proposed method is applicable to normal urban driving speed. As we have mentioned in Section III B, the high frame rate of ground-view camera guarantees the feasibility of proposed method. With the increase in vehicle speed, as the translation between two consecutive images increases, a higher frame rate might be needed to make sure camera$_i$ planar motion $\Delta\mathbf{T}_g^{k,\,k+1}$ can be accurately estimated. However, in practice, it is reasonable to limit the vehicle speed when the online calibration is conducted, so we think the performance of proposed method on different vehicle speeds is satisfactory.

B. Field Test

In the field test, we used a small wheeled robot, RoboCar 1/10 (ZMP Inc.), as shown in Fig. 7. Two industrial cameras (MQ013MG-ON, XIMEA Inc.) were installed on the robot. The camera to be calibrated faced forward with a frame rate of 18 Hz (resolution: 1280×1024 pixels), and another camera faced the ground with a frame rate of 180 Hz (resolution: 800×600 pixels); the height of the ground-view camera from the ground was 0.135 m. The ground truth of the extrinsic parameters of the two cameras was obtained using the OptiTrack system [27]. Three different image sequences were collected using the same camera configuration when the robot traveled on a regular road with a mixed trajectory composed of

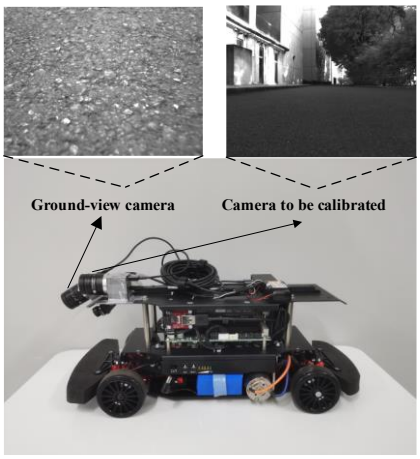

Fig. 7. Mobile robot used in field test and sample images captured by cameras.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

straight motions and several turns. The performance of SensorX2car method and proposed method was also compared in the field test. For the proposed method, similar to the simulation test, these image sequences were subdivided into multiple subsequences, and the average calibration results were used as the calibration result of each image sequence. The length of each sub-sequence was the same as that used in the simulations. Another level of extrinsic parameters' initial errors with 3° error in rotation and 0.02 m error in translation was also used for proposed method for more comprehensive evaluation.

Table III lists the field test results of two methods and all three angles of $\mathbf{R}_{va}$ are shown. From the table, we can see that for proposed method with error level 1, although pitch and roll angle calibration error in Sequence 1 and roll angle calibration error in Sequence 3 increase to above 1°, most of the calibration errors are still within 1°. Significantly, the calibration errors of the yaw angle, which have the greatest influence on the accuracy of obstacle detection in vehicular scenarios, are all within 0.55°. Compared with SensorX2car method, in the aspect of yaw and pitch angle calibration accuracy, proposed method with error level 1 is better and more stable except yaw angle calibration accuracy in sequence 3; while in the aspect of roll angle calibration accuracy, SensorX2car method is better. For proposed method with error level 2, the calibration errors increase: as Table III shows, most are above 1° and compared with SensorX2car method, it only has higher accuracy in pitch angle calibration. Based on the above analysis, the overall performance of the proposed method in the field test can meet the target accuracy and is comparable with SensorX2car method; however, when there are significant errors in the initial extrinsic parameters, the calibration errors increase to above 1°.

C. Discussion

From the results of the simulation and field tests, we can see that the overall performance of the proposed algorithm can achieve target accuracy. Notably, the field test results were inferior to the simulation test results. This is because the mobile wheeled robot used in the field test is considerably smaller than a normal-sized car. Therefore, when it travels on a normal road, it suffers from much more vibration than an ordinary car, which affects the assumption of extrinsic parameters invariance. As a result, the final calibration results will be like the "average" of changing parameters, thus the calibration accuracy of the algorithm is decreased compared with that in the simulation test and has worse performance with the increase of initial errors. This inference is verified through a simulation test. When ego car was running on a normal road, to simulate the situation of motion on a bumpy road, different amplitudes of stochastic

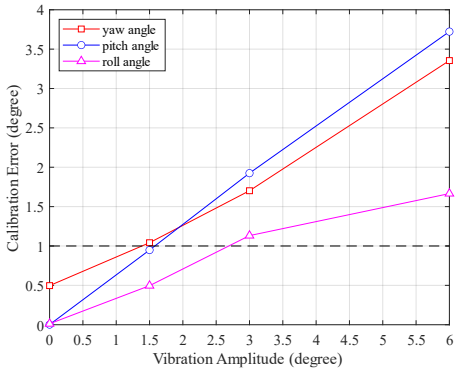

Fig. 8. Orientation calibration errors under different amplitudes of camera vibration during vehicle motion. All of the calibration errors are absolute values.

vibration were added to both cameras' orientation relative to vehicle, extrinsic parameters' initial errors with 5° error in rotation and 0.03 m error in translation were used and the test results are shown in Fig. 8. From the test results, we can see that with the increase of vibration amplitude, the calibration errors of camera orientation also increase, which verifies that strong vibration degrades the performance of proposed method; however, calibration error can still be limited to approximately 1° when vibration is not strong, as the test results of 1.5° vibration amplitude show, which reveals the robustness of the proposed method when used on a bumpy road and the same conclusion can also be drawn from field test results with an initial rotation error of 3°. Although the calibration accuracy in the field test decreased when the initial rotation error reached 5°, in practice, when a vehicle runs on a regular road in a stable status, it does not suffer as much vibration as the scale-down car does in field test. This can also be observed in the supplementary video, where a simulation environment video recorded in Simulation Test part and a KITTI [28] video are compared with that recorded in the field test environment, which exhibits evident vibration compared to the former two. Besides, the errors in the initial extrinsic parameters can also be limited with regular online calibration to ensure that calibration accuracy is guaranteed.

Overall, compared with SensorX2car like method based on VPs detection, proposed method has weaker limitations in the type of sensor to be calibrated and it does not require that the sensor to be calibrated be installed in a position where road direction can be observed and does not consider the road direction as vehicle's direction. Moreover, its accuracy is better than that of SensorX2car method when the vehicle is traveling on a regular road in a stable status and comparable with it even

TABLE III. FIELD TEST RESULTS OF TWO DIFFERENT METHODS

		SensorX2car [24]			Proposed method (error level 1)[b]			Proposed method (error level 2)[b]		
		Yaw	*Pitch*	*Roll*	*Yaw*	*Pitch*	*Roll*	*Yaw*	*Pitch*	*Roll*
Sequence 1	Result	0.778	-1.258	0.247	0.712	2.102	1.274	3.615	1.874	-2.077
	Error	0.362	-1.910	0.789	0.296	1.450	1.816	3.198	1.222	-1.536
Sequence 2	Result	-9.091	-0.137	0.166	0.092	0.139	0.361	0.700	-0.078	-0.477
	Error	-9.508	-0.789	0.708	-0.324	-0.513	0.903	0.284	-0.730	0.064
Sequence 3	Result	0.747	-0.576	0.027	-0.132	0.581	0.801	2.862	-0.455	0.881
	Error	0.331	-1.228	0.568	-0.549	-0.071	1.343	2.446	-1.107	1.423

[b]Error level 1 represents 3° initial error in rotation and 0.02m initial error in translation, error level 2 represents 5° initial error in rotation and 0.03m initial error in translation.

on bumpy road. Therefore, we can say that proposed method is superior to SensorX2car method and can achieve the target accuracy. As mentioned earlier, we address only part of the complete calibration problem, as a coarse initial value of a camera's extrinsic parameters is expected. However, we believe this simplification is reasonable and valid for practical application. Our target is to free the camera orientation from the disturbance of various factors, such as payload variety and significant temperature variations, and guarantee its accuracy during a vehicle life cycle, not to completely replace offline calibration in the production line. With the proposed algorithm embedded in the vehicle, the extrinsic parameters of the camera obtained during the vehicle production process can be used as initial values to calibrate the camera orientation online. Besides, the calibration results of methods like SensorX2car can also be used as the initial values of proposed method to achieve better accuracy.

V. Conclusion

This study proposes a novel online camera orientation calibration method that is independent of artificial features for onboard cameras on vehicles. A high-speed ground-view camera was introduced into the framework as an intermediary to aid the calibration of target camera's orientation with respect to vehicle. By using a semi-dense approach to efficiently analyze the road surface and conducting nonlinear optimization based on the extrinsic parameter invariance constraint and the vehicle nonholonomic constraint to calibrate the camera orientation, we can limit the calibration error to approximately 1°, even under poor road conditions.

The proposed method can be used when there is a coarse value of extrinsic parameters, and it is also robust under bumpy road. Regular execution of the method can deal with slowly changing extrinsic parameters. In the future, this method can be used in automated driving systems equipped with other types of sensors that can estimate their own motions. With the development of sensor motion estimation algorithms, SOTA methods can be introduced to our method to increase its accuracy. Currently, the method requires about several minutes in a normal laptop to get results owing to a prototyping implementation. Although it is acceptable due to the stability of camera extrinsic parameters in a short period, we plan to accelerate it through parallel implementation in the future.

References

[1] I. Cvišić, I. Marković, and I. Petrović, "Recalibrating the KITTI dataset camera setup for improved odometry accuracy," in *Proc. Euro. Conf. Mob. Robots*, 2021, pp. 1–6.

[2] T. Dang, C. Hoffmann, and C. Stiller, "Continuous stereo self-calibration by camera parameter tracking," *IEEE Trans. Img. Process.*, vol. 18, no. 7, pp. 1536–1550, Jul. 2009.

[3] J. H. Lee and D.-W. Lee, "A hough-space-based automatic online calibration method for a side-rear-view monitoring system," *Sensors*, vol. 20, no. 12, pp. 3407, Jun. 2020.

[4] M. B. de Paula, C. R. Jung, and L. G. da Silveira Jr, "Automatic on-the-fly extrinsic camera calibration of onboard vehicular cameras," *Exp. Syst. with App.*, vol. 41, no. 4, pp. 1997-2007, 2014.

[5] K. Zhao, U. Iurgel, M. Meuter, and J. Pauli, "An automatic online camera calibration system for vehicular applications," in *Proc. Int. IEEE Conf. Intel. Transp. Syst.*, 2014, pp. 1490–1492.

[6] K. Choi, H. Jung, and J. Suhr, "Automatic calibration of an around view monitor system exploiting lane markings," *Sensors*, vol. 18, no. 9, pp. 2956, Sep. 2018.

[7] M. Bellusci and M. Matteucci, "Advances in real-time online vehicle camera calibration via road line markings parallelism enforcement," in *Proc. IEEE Intel. Veh. Symp.*, 2022, pp. 1511–1516.

[8] R. Itu, D. Borza, and R. Danescu, "Automatic extrinsic camera parameters calibration using convolutional neural networks," in *Proc. IEEE Int. Conf. on Intel. Comp. Comm. Processing*, 2017, pp. 273–278.

[9] J. Jang, Y. Jo, M. Shin, and J. Paik, "Camera orientation estimation using motion-based vanishing point detection for advanced driver- assistance systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 10, pp. 6286–6296, Oct. 2021.

[10] J. Lee, M. Sung, H. Lee, and J. Kim, "Neural geometric parser for single image camera calibration," in *Proc. Euro. Conf. Comp. Vis.*, 2020, pp. 541–557.

[11] M. Miksch, B. Yang, and K. Zimmermann, "Automatic extrinsic camera self-calibration based on homography and epipolar geometry," in *Proc. IEEE Intel. Veh. Symp.*, 2010, pp. 832–839.

[12] M. Knorr, W. Niehsen, and C. Stiller, "Online extrinsic multi-camera calibration using ground plane induced homographies," in *Proc. Intel. Veh. Symp.*, 2013, pp. 236–241.

[13] J. Tan, X. An, X. Xin, and H. He, "Automatic extrinsic calibration for an onboard camera," in *Proc. Chinese Automat. Congr*, 2013, pp. 340–343.

[14] R. G. Mueller, P. Burger, and H.-J. Wuensche, "Continuous stereo self-calibration on planar roads," in *Proc. IEEE Intel. Veh. Symp.*, 2018, pp. 1755-1760.

[15] J. Zienkiewicz, R. Lukierski, and A. Davison, "Dense, auto-calibrating visual odometry from a downward-looking camera," in *Proc. British Mach. Vis. Conf.*, 2013, pp. 1-11.

[16] M. Hirano, T. Senoo, N. Kishi, and M. Ishikawa, "Virtual inverse perspective mapping for simultaneous pose and motion estimation," *arXiv preprint arXiv:2303.05192*, 2023.

[17] Assessment Protocol—Vulnerable Road User Protection, *European New Car Assessment Programme (Euro NCAP)*, Nov. 2022.

[18] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, "High-accuracy subpixel image registration based on phase-only correlation," *IEICE Trans. Fundam. Electron. Comm. Comp. Sci.*, vol. E86-A, no. 8, pp. 1925–1934, Aug. 2003.

[19] G. Dubbelman and B. Browning, "COP-SLAM: Closed-form online pose-chain optimization for visual SLAM," *IEEE Trans. Robot.*, vol. 31, no. 5, pp. 1194–1213, Oct. 2015.

[20] G. Grisetti, R. Kümmerle, H. Strasdat, and K. Konolige, "g2o: A general framework for (hyper) graph optimization," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2011, pp. 9–13.

[21] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Trans. Pat. Anal. Mach. Intel.*, vol. 13, no. 04, pp. 376–380, Apr. 1991.

[22] J. Zienkiewicz and A. Davison, "Extrinsics autocalibration for dense planar visual odometry," *J. of Field Robot.*, vol. 32, no. 5, pp. 803-825, Aug. 2015.

[23] P. Lébraly, "Calibration of non-overlapping cameras—Application to vision-based robotics," in *Proc. British Mach. Vis. Conf.*, 2010, pp. 1-12.

[24] G. Yan, Z. Luo, Z. Liu, and Y. Li, "SensorX2car: Sensors-to-car calibration for autonomous driving in road scenarios," *arXiv preprint arXiv:2301.07279*, 2023.

[25] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proc. Conf. Robot Learn.*, 2017, pp. 1–16.

[26] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: An open-source SLAM system for monocular, stereo, and RGB-D cameras," *IEEE Trans. Robot.*, vol. 33, no. 5, pp. 1255–1262, Oct. 2017.

[27] NaturalPoint, Inc. (2023, Mar. 10) *OptiTrack System*. [Online] Available: https://optitrack.com/.

[28] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comp. Vis. Pat. Recog.*, 2012, pp. 3354–3361.