

A Generic Vehicle-to-Sensor Calibration Framework

Sumin Hu Youngmin Yoo Jeeseong Kim Changsoo Lim Doohyun Cho* Bongnam Kang
 StradVision

Abstract

In autonomous driving systems, online vehicle-to-sensor (v2s) calibration is a critical component for ensuring safe perception-based control. Since sensor pose may shift during the life-time of a vehicle, online calibration is essential to maintain safe driving conditions. To this end, this paper introduces *Epipoles as a 3D Directional Compass (E3DC)*, a sensor-agnostic v2s online calibration method. Leveraging the nonholonomic nature of vehicles, a hand-eye constraint between the vehicle and the sensor naturally emerges. Consequently, we require only the sensor’s data to determine the v2s extrinsic rotation. More specifically, since we only require sensor odometry estimates to perform v2s calibration, E3DC can leverage off-the-shelf odometry estimation pipelines. This offers vast flexibility and wide applicability as the odometry estimation pipeline can be tailored to the specific sensor type and driving environment. We demonstrate that our method is robust and achieves state-of-the-art performance on both the KITTI dataset and a new dataset, which will be made publicly available. To the best of our knowledge, this is the first v2s calibration dataset for autonomous driving scenarios.

1. Introduction

The pursuit of fully autonomous vehicles has been a central focus of research for decades, leading to substantial advancements in critical areas such as object detection, lane detection, and depth estimation [17, 40, 43]. Concurrently, the integration of increasingly sophisticated sensor arrays into self-driving cars has become essential for enhancing passenger safety, leveraging the maturity of vision algorithms in these domains [5, 6]. A key aspect of ensuring the reliability of these systems is implementing online sensor calibration as a fail-safe mechanism, crucial for maintaining the accuracy of sensor measurements. In particular, online sensor-to-sensor (s2s) calibration has gained significant research attention [3, 9, 10, 15, 19, 22, 23, 25, 26, 35, 37–

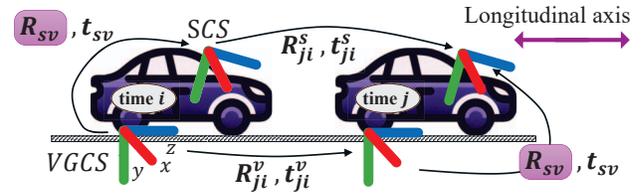


Figure 1. **Illustration of the online vehicle-to-sensor extrinsic rotation calibration problem.** We aim to determine the extrinsic rotation, R_{sv} between the vehicle and sensor as the vehicle moves over time, e.g., time i to j . Refer to Sec. 3 for more details.

39, 41, 42, 44], as it enables the precise alignment of data from diverse sensors, thereby enhancing the reliability of perception tasks.

While past studies do exist, online vehicle-to-sensor (v2s) calibration has received comparatively little attention [8, 12–14, 16, 27, 28, 31, 33, 36]. Accurate online v2s calibration is critical for ensuring that detected objects are correctly localized relative to the vehicle, which is essential for the safe control of autonomous systems, including the reliable deployment of vehicle safety features such as autonomous emergency braking.

Despite the importance of v2s calibration, current public datasets like KITTI and a2d2 [5, 6] do not offer sufficiently precise v2s calibration results suitable for evaluating v2s calibration algorithms. For instance, the KITTI dataset assumes alignment between the GPS/IMU sensor at the rear axle and the vehicle’s coordinate system, while a2d2 dataset presumes the accuracy of the main LiDAR’s extrinsic calibration relative to the vehicle frame. Additionally, accurately determining an object’s 3D position relative to the ground on which the vehicle stands is essential for safe control. However, factors such as load distribution shifts can alter the vehicle’s orientation, thereby affecting sensor alignment. To address these issues, we introduce a novel dataset with accurate v2s extrinsics. This dataset includes six cameras, providing 360-degree surround view, offering a foundation for research in v2s calibration.

As previously noted by Su *et al.* [27], rotation misalignment tends to have a more pernicious effect compared to translation misalignment, which is why rotation calibration

*Work completed while employed at StradVision.

is often prioritized. This priority is further justified by the fact that rotation is more susceptible to changes due to factors such as collisions, replacements of vehicle parts, or shifts in load distribution. These vulnerabilities necessitate frequent recalibration to ensure accurate sensor alignment. Consequently, much of the existing research, including our own, has focused on developing robust methods for extrinsic rotation estimation, as depicted in Fig. 1, to effectively address these challenges.

Reflecting the need for precise rotation calibration, prior studies have used various methods. A classic approach involves utilizing lane lines to determine the Longitudinal Vanishing Point (LVP) [8, 12, 31, 36], which represents the vehicle’s forward direction w.r.t. the sensor. Some rely on feature detection on the ground plane to derive the ground’s normal vector via homography decomposition [16, 28, 33]. A drawback of using lane lines or ground features is that these approaches are effective only in environments where the relevant features are abundant. Furthermore, since these methods are specific to a sensor and its setup, they are limited in flexibility and difficult to adapt to different configurations.

To overcome these limitations, we propose *Epipoles as a 3D Directional Compass (E3DC)*, a framework that is adaptable to different sensors and environments. By decoupling the calibration process from sensor-specific features, E3DC enables v2s rotation calibration using any sensor, as long as the odometry is obtained. We have tested E3DC for cameras on the KITTI dataset [5] and the new custom dataset, showcasing state-of-the-art performance in all surround-view directions. While our primary focus is on cameras in this paper, we also present LiDAR results in the supplementary material.

2. Related Works

Approaches to v2s calibration vary widely, encompassing different assumptions, prior knowledge, favorable environments, and auxiliary sensor requirements. We summarize the dependencies of each method in Tab. 1.

Lane-based methods, such as S2XV [36], utilize Line Segment Detection (LSD) [32] to identify lines, which are then processed by a deep learning network to detect vanishing points (VPs) of the road and the horizon line (HL). Other approaches also leverage lane markings to determine the VP [8, 13, 31]. However, a fundamental limitation of these methods is that VPs represent the alignment of the camera with the road, *i.e.*, world, rather than with the vehicle itself. Consequently, Kalman filters [8, 13] or heuristic thresholds [31, 36] are required to confirm vehicle-road alignment, which is suboptimal. Additionally, the effectiveness of these methods is limited to environments with clearly defined straight lines, and some methods even require known lane widths to estimate the camera’s roll an-

Method	Auxiliary Sensor	Sensor Installation	Environment	Prior Knowledge
Miksch <i>et al.</i> 2010 [16]	⚠	⚠	⚠	⚠
Tan <i>et al.</i> 2013 [28]	⚠	⚠	⚠	⚠
Westerhoff <i>et al.</i> 2016 [33]	⚠	⚠	✔	⚠
Tummala <i>et al.</i> 2019 [31]	✔	⚠	⚠	⚠
Ouyang <i>et al.</i> 2020 [21]	✔	✔	⚠	✔
Lee <i>et al.</i> 2020 [13]	✔	⚠	⚠	⚠
Lee and Lee 2020 [12]	✔	⚠	✔	⚠
Jang <i>et al.</i> 2021 [8]	✔	⚠	⚠	⚠
Li <i>et al.</i> 2023 [14]	⚠	⚠	✔	✔
Su <i>et al.</i> 2023 [27]	⚠	⚠	✔	✔
SX2V (Yan <i>et al.</i> 2024) [36]	✔	⚠	⚠	✔
E3DC (ours)	✔	✔	✔	✔

Table 1. **Comparison of each method’s dependencies.** Green shield (✔) indicates that the method is *independent* of this factor, demonstrating robustness and flexibility. Yellow warning (⚠) indicates that the method is *dependent* of this factor, limiting its applicability.

gle [8].

Homography-based methods [16, 33] utilize road features to estimate a homography matrix, which can be decomposed to derive the road surface normal - another form of the HL. Despite showing promise, these methods necessitate environments with globally flat road surfaces. Additionally, road surfaces are often not ideal for tracking features, posing another significant limitation. Moreover, these approaches typically depend on precise measurements of camera height and additional vehicle odometry data.

While using lanes and road surfaces are valid and proven methods for v2s calibration, these methods inherently require sacrificing a portion of the camera’s field of view to focus on the road. This trade-off can limit the system’s ability to monitor potentially more critical elements, such as nearby vehicles or pedestrians. Additionally, these methods are generally unsuitable for side-view cameras, as they rely on front or rear-facing orientations to function effectively. In contrast, the proposed method E3DC is adaptable across all viewing directions, provided that the target sensor’s odometry data is available.

E3DC circumvents the complexities associated with additional sensor inputs required by other approaches. For instance, the use of high-speed cameras aimed at the ground may not be practical for widespread implementation in autonomous vehicles [27]. Additionally, methods that rely on vehicle odometry can have their reliability compromised by physical changes in the wheels, such as wear or replacement. Therefore, E3DC focuses on leveraging only the single target sensor for v2s calibration, eliminating the need for temporal sensor fusion and simplifying the overall process.

3. Proposed Method

The overview of our method is shown in Fig. 2. While we eventually show our method is sensor-agnostic, we first formulate the solution w.r.t. the camera sensor.

Coordinate Systems. As shown in Fig. 1, we work with

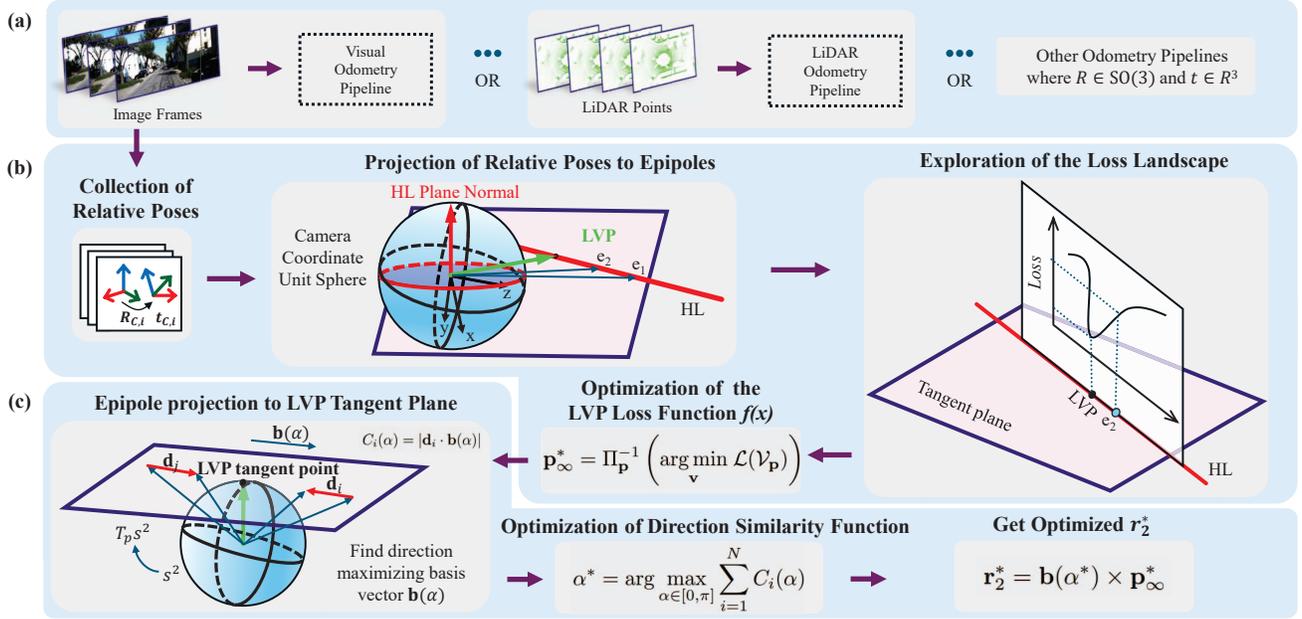


Figure 2. **Overview of E3DC.** (a) The odometry estimation pipeline for a target sensor and environment. (b) Robust Longitudinal Vanishing Point (LVP) estimation pipeline. (c) Robust Horizon Line (HL) estimation pipeline.

two coordinate systems: the Vehicle Ground Coordinate System (VGCS) and the Sensor Coordinate System (SCS). As the vehicle moves, 3D points in time frame i are transformed to points in time frame j for each coordinate system in the following manner:

$$\mathbf{X}_j^v = \mathbf{R}_{ji}^v \mathbf{X}_i^v + \mathbf{t}_{ji}^v, \quad (1)$$

$$\mathbf{X}_j^s = \mathbf{R}_{ji}^s \mathbf{X}_i^s + \mathbf{t}_{ji}^s. \quad (2)$$

The VGCS has its Z-axis aligned with the vehicle's longitudinal axis and the XZ plane parallel to the ground. The term VGCS was introduced to emphasize that this coordinate system remains unaffected by changes in the vehicle body's orientation, such as those caused by shifts in load distribution. Thus, the transformation between the VGCS and the SCS can be described by the rotation matrix \mathbf{R}_{sv} and translation vector \mathbf{t}_{sv} , satisfying the equation:

$$\mathbf{X}^s = \mathbf{R}_{sv} \mathbf{X}^v + \mathbf{t}_{sv}, \quad (3)$$

where \mathbf{X}^s and \mathbf{X}^v physically represent the same 3D point in the SCS and the VGCS, respectively. Therefore, we have the relationships

$$\mathbf{R}_{ji}^s = \mathbf{R}_{sv} \mathbf{R}_{ji}^v \mathbf{R}_{sv}^T, \quad (4)$$

$$\mathbf{t}_{ji}^s = \mathbf{R}_{sv} (\mathbf{t}_{ji}^v (\mathbf{I} - \mathbf{R}_{ji}^v) \mathbf{R}_{sv}^T \mathbf{t}_{sv}). \quad (5)$$

The rotation matrix \mathbf{R}_{sv} can be further expressed in Euler angles and column vectors as follows:

$$\mathbf{R}_{sv} = \mathbf{R}_z(\phi) \mathbf{R}_x(\theta) \mathbf{R}_y(\psi) \quad (6)$$

$$= [\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{r}_3], \quad (7)$$

where roll is ϕ , pitch is θ , yaw is ψ , and \mathbf{r}_i is the i th column vector of the rotation matrix.

Longitudinal Vanishing Point (LVP). The LVP denotes the image point of \mathbf{Z}_∞^v , which is the point at infinity in the Z-axis of the VGCS. Its formulation is as follows:

$$\mathbf{p}_\infty^s = \mathbf{K} [\mathbf{R}_{sv} \quad \mathbf{t}_{sv}] \mathbf{Z}_\infty^v = \mathbf{K} \mathbf{r}_3, \quad (8)$$

where \mathbf{p}_∞^s is the LVP in the SCS, \mathbf{K} is the intrinsic camera matrix, and $\mathbf{Z}_\infty^v = [0 \ 0 \ 1 \ 0]^T$ in the homogeneous coordinate system. Knowing the LVP, we can retrieve \mathbf{r}_3 as

$$\mathbf{r}_3 = \frac{\mathbf{K}^{-1} \mathbf{p}_\infty^s}{\|\mathbf{K}^{-1} \mathbf{p}_\infty^s\|}, \quad (9)$$

where $\|\cdot\|$ is the L2 norm.

Horizon Line (HL). The Horizon Line (HL) is the image projection of the line at infinity on the ground plane, *i.e.*, the XZ plane of the VGCS. The formulation of the HL can be obtained by connecting two points at infinity in the XZ plane from the VGCS. Say we have points at infinity in the X and Z axes directions in the VGCS, *i.e.*, $\mathbf{X}_\infty^v = [1 \ 0 \ 0 \ 0]$ and $\mathbf{Z}_\infty^v = [0 \ 0 \ 1 \ 0]$, projected to the image plane:

$$\mathbf{p}_x = \mathbf{K} [\mathbf{R}_{sv} \quad \mathbf{t}_{sv}] \mathbf{X}_\infty^v = \mathbf{K} \mathbf{r}_1, \quad (10)$$

$$\mathbf{p}_z = \mathbf{K} [\mathbf{R}_{sv} \quad \mathbf{t}_{sv}] \mathbf{Z}_\infty^v = \mathbf{K} \mathbf{r}_3. \quad (11)$$

By taking the cross product of the two points we get the HL:

$$\mathbf{l}_H = \mathbf{K} \mathbf{r}_1 \times \mathbf{K} \mathbf{r}_3 = \mathbf{K}^* (\mathbf{r}_1 \times \mathbf{r}_3) \quad (12)$$

$$= \mathbf{K}^* \mathbf{r}_2, \quad (13)$$

where $\mathbf{K}^* = \det(\mathbf{K})\mathbf{K}^{-T}$ is the adjoint of \mathbf{K} . Therefore, knowing the HL, we can retrieve \mathbf{r}_2 as

$$\mathbf{r}_2 = \frac{\mathbf{K}^T \mathbf{l}_H}{\|\mathbf{K}^T \mathbf{l}_H\|}, \quad (14)$$

where $\|\cdot\|$ is the L2 norm.

Obtaining the Extrinsic Rotation \mathbf{R}_{sv} . Given that the LVP and the HL is obtained, the extrinsic rotation \mathbf{R}_{sv} can be determined by finding \mathbf{r}_1 as

$$\mathbf{r}_1 = \mathbf{r}_2 \times \mathbf{r}_3 \quad (15)$$

and concatenating the three column vectors. Consequently, Eqs. (8), (13) and (15) demonstrates that v2s calibration can be decoupled to two distinctive problems: Identifying the LVP and the HL.

3.1. LVP and HL are Epipoles

With readily available relative pose estimation methods, we obtain the camera's relative pose: \mathbf{R}_{ji}^s and \mathbf{t}_{ji}^s . The homogeneous coordinate representation of 2D epipoles $\mathbf{e}_k^{s,I} \in \mathbb{R}^3$ on the image plane I are defined as:

$$\mathbf{e}_k^{s,I} := \begin{cases} \mathbf{K}\mathbf{R}_{ji}^s \mathbf{t}_{ji}^s & \text{for } k = 1, \\ \mathbf{K}\mathbf{t}_{ji}^s & \text{for } k = 2. \end{cases} \quad (16)$$

We extensively exploit these epipoles to perform v2s calibration. First, we note that in previous studies such as [1,2,7], a key observation was made: In scenarios involving planar motion, epipoles are invariably located on the HL. In other words, for a vehicle traversing in planar motion, the HL is the pencil of epipoles:

$$\mathbf{l}_H = \{ \mathbf{e}^{s,I} \mid \mathbf{R}_{ji}^s = \exp(\gamma[\boldsymbol{\pi}_g^s]_{\times}), \mathbf{t}_{ji}^s \perp \boldsymbol{\pi}_g^s \}, \quad (17)$$

where γ is the rotation angle, $\boldsymbol{\pi}_g^s$ denotes the ground plane normal in the SCS.

Additionally, with the LVP and HL represented as Eq. (8) and Eq. (13), we easily confirm that the LVP is on the HL since $\mathbf{l}_H^T \mathbf{p}_\infty = 0$.

As epipoles point to the LVP and HL, they act as directional cues for v2s calibration. Thus, the methods is named *Epipoles as a 3D Directional Compass (E3DC)* to reflect their role as a 3D guiding tool for calibration.

3.2. Robust LVP Estimation

With Eqs. (4) and (5), and \mathbf{K} , we can model the epipoles with the spherical projection model [30] in the VGCS:

$$\mathbf{e}_1^v := \mathbf{K}^{-1} \mathbf{e}_1^{c,I} = \mathbf{t}_{ij}^v + (\mathbf{R}_{ij}^v - \mathbf{I}) \mathbf{t}_{vs}, \quad (18)$$

$$\mathbf{e}_2^v := \mathbf{K}^{-1} \mathbf{e}_2^{c,I} = \mathbf{t}_{ji}^v + (\mathbf{R}_{ji}^v - \mathbf{I}) \mathbf{t}_{vs}, \quad (19)$$

where the detailed derivations are provided in the supplementary material.

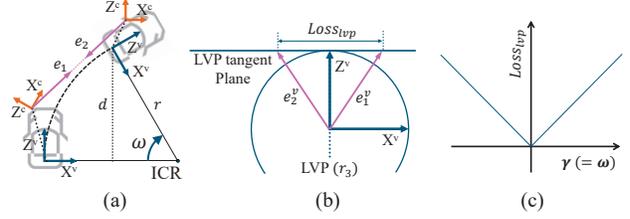


Figure 3. **Ackermann guarantees convexity for LVP loss function.** (a) Ackermann model (b) $Loss_{lvp}$ is the distance between epipoles projected onto a LVP tangent plane. (c) Loss landscape of $Loss_{lvp}$ as the angle-axis magnitude γ from the camera changes.

Additionally, as depicted in Fig. 3(a) and explained in [34], the nonholonomic constraints inherent in the Ackermann steering geometry restricts a vehicle's movement to a circular trajectory with the axis of rotation located at the Instantaneous Circle of Rotation (ICR):

$$\mathbf{R}_{ji}^v = \begin{bmatrix} \cos(\omega) & 0 & \sin(\omega) \\ 0 & 1 & 0 \\ -\sin(\omega) & 0 & \cos(\omega) \end{bmatrix}, \mathbf{t}_{ji}^v = \frac{d}{\sin(\omega)} \begin{bmatrix} 1 - \cos(\omega) \\ 0 \\ \sin(\omega) \end{bmatrix}, \quad (20)$$

where d is the displacement towards the Z^v direction. It is easily verifiable that $\mathbf{t}_{ji}^v = [0 \ 0 \ 1]^T \iff \mathbf{R}_{ji}^v = \mathbf{I} \iff \mathbf{R}_{ji}^s = \mathbf{I}$ when $\omega = 0$. In this situation, Eq. (16) greatly simplify as follows:

$$\mathbf{e}_1^{c,I} = \mathbf{e}_2^{c,I} = \mathbf{K}\mathbf{R}_{sv} \mathbf{t}_{ji}^v \quad (21)$$

$$= \mathbf{K}\mathbf{R}_{sv} [0 \ 0 \ 1]^T \quad (22)$$

$$= \mathbf{K}\mathbf{r}_3. \quad (23)$$

Notice that the definition of LVP in Eq. (8) matches the epipole notation in Eq. (23) for straight vehicle motion. This realization underscores that the nonholonomic motion inherently ensures a hand-eye constraint between the vehicle and camera.

Furthermore, an analysis of the epipoles projected to a tangent plane by the projection function $\Pi_{\mathbf{r}_3} : S^2 \rightarrow T_{\mathbf{r}_3} S^2$, where \mathbf{r}_3 is the ideal LVP, showcases that Ackermann steering geometry guarantees convexity in the LVP loss function. The projection function $\Pi_{\mathbf{r}_3}$ maps an epipole \mathbf{e} on the unit sphere to a point \mathbf{v} in the tangent plane. Substituting Eqs. (4) and (5) in Eqs. (18) and (19) and utilizing small angle approximations of ω , the x-axis coordinate of \mathbf{v}_1 and \mathbf{v}_2 reduces to

$$v_{1,x} \approx \omega \left(\frac{1}{2} + \frac{t_{vs3}}{d} \right) + O(\omega^2), \quad (24)$$

$$v_{2,x} \approx \omega \left(-\frac{1}{2} + \frac{t_{vs3}}{d} \right) + O(\omega^2), \quad (25)$$

where t_{vs3} is signed displacement of the camera in the longitudinal axis w.r.t. to VGCS and $O(\omega^2)$ denote higher order terms. We define a geodesic loss

$$Loss(\mathbf{v}_1, \mathbf{v}_2) = \|\mathbf{v}_1 - \mathbf{v}_2\|, \quad (26)$$

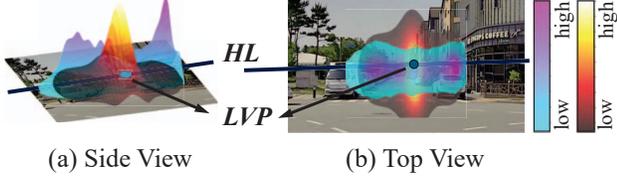


Figure 4. **Illustration of geodesic loss and probability density function.** (a) Side view. (b) Top view. The bluish-purple color illustrates the loss function $\mathcal{L}(\mathcal{V}_p)$ at epipoles mapped to the image. It increases as it gets further away from the LVP. The reddish-yellow color illustrates the PDF of epipoles sampled from a typical driving scene. They are highly concentrated near the LVP.

with the ideal scenario resulting in

$$Loss_{lvp}(\omega) = |v_{1,x} - v_{2,x}| = |\omega|. \quad (27)$$

Noting that trace is invariant under circular shifts, *i.e.*, $\text{Tr}(\mathbf{R}_{j_i}^s) = \text{Tr}(\mathbf{R}_{sv} \mathbf{R}_{j_i}^v \mathbf{R}_{sv}^T) = \text{Tr}(\mathbf{R}_{j_i}^v \mathbf{R}_{sv}^T \mathbf{R}_{sv}) = \text{Tr}(\mathbf{R}_{j_i}^v)$, we know that ω can be retrieved from $\mathbf{R}_{j_i}^s$ with the axis-angle rotation magnitude $\gamma = \cos^{-1}\left(\frac{\text{Tr}(\mathbf{R}_{j_i}^s) - 1}{2}\right)$. Therefore, Fig. 3(c) depicts how Ackermann steering principle manifests a convex loss function $Loss(\mathbf{v}_1, \mathbf{v}_2)$, which can be optimized to obtain the optimized LVP

$$\mathbf{p}_\infty^* = \Pi_{\mathbf{p}}^{-1}\left(\arg \min_{\mathbf{v}} \mathcal{L}(\mathcal{V}_p)\right), \quad (28)$$

where \mathcal{V}_p is the set of epipole pairs and $\mathcal{L}(\mathcal{V}_p)$ is the loss function created from epipole and loss pairs, *i.e.*, $\left\{\left(\frac{\mathbf{v}_{1,i} + \mathbf{v}_{2,i}}{2}\right), Loss(\mathbf{v}_{1,i}, \mathbf{v}_{2,i})\right\}$.

As illustrated by the high Probability Density Function (PDF) near the LVP in Fig. 4, the nonholonomic constraint of a ground vehicle naturally causes it to traverse more frequently in the LVP direction and thus allows for sufficient samples of epipoles and loss pairs near the LVP.

3.3. Robust HL Estimation

As shown in Sec. 3.1, the LVP is a point on the HL, which means that solving for Eq. (28) provides a highly confident reference point for HL estimation. Additionally, since epipoles lie on the HL for planar motion, the problem simplifies to finding the best direction that aligns the epipoles. The directional alignment is evaluated by the Unnormalized Absolute Cosine Similarity (UACS):

$$C_i(\alpha) = |\mathbf{d}_i \cdot \mathbf{b}(\alpha)|, \quad (29)$$

where the unit vector $\mathbf{b}(\alpha) = \mathbf{R}(\alpha)\mathbf{b}_0$ is formed by rotating an initial basis unit vectors \mathbf{b}_0 rotated by $\mathbf{R}(\alpha) = \exp(\alpha[\mathbf{p}_\infty^*]_x)$ in the tangent plane, given $\mathbf{p}_\infty^* \in \mathbb{R}^3$. The goal is to find angle α^* that maximizes the UACS across all epipole pairs:

$$\alpha^* = \arg \max_{\alpha \in [0, \pi]} \sum_{i=1}^N C_i(\alpha). \quad (30)$$

The optimal rotation axis is then determined as

$$\mathbf{r}_2^* = \mathbf{b}(\alpha^*) \times \mathbf{p}_\infty^*, \quad (31)$$

where $\mathbf{b}(\alpha^*) \in \mathbb{R}^3$.

3.4. LVP and HL Estimation Pipeline

With N number of elements per batch, the set of relative poses in batch m is

$$\mathcal{T}_m = \left\{ \left\{ \mathbf{R}_{i+1,i}^s, \mathbf{t}_{i+1,i}^s \right\} \mid i = 0, 1, \dots, N \right\}, \quad (32)$$

where i denote the time. At the first batch, Eq. (28) requires a reference point \mathbf{p} to project the epipoles to $\mathbf{v} = \Pi_{\mathbf{p}}(\mathbf{e})$. Thus, a unit average epipole direction vector is used instead. After optimizing Eq. (28), a better LVP estimate of $\mathbf{p}_{m=1}$ is used to project the points and perform HL optimization. Then, the process repeats with N more new elements of relative poses added to the set: $\mathcal{T}_{1:m} = \mathcal{T}_{1:m-1} + \mathcal{T}_m$.

3.5. Sensor Independent E3DC Formulation

The construction of identifying the HL and LVP allows the solution to be easily expanded to other sensor modules such as LiDAR. This is evident since the \mathbf{K} matrix used in Eq. (9) and Eq. (14) is essentially redundant when the HL and LVP are represented in the unit sphere as shown in Fig. 2. This signifies the redundancy of \mathbf{K} in obtaining the extrinsic rotation and, therefore, proves that this method is truly independent of sensors.

4. Results

4.1. Sensitivity Analysis

The v2s extrinsic rotation (\mathbf{R}_e) is a parameter that can be influenced by factors such as where the passengers are seated. As shown in Fig. 5, depending on where passengers take seat, the extrinsic rotation changes, with a maximum difference of 0.40 degrees in pitch and 0.29 degrees in roll. While s2s calibrations are not affected by these difference, since s2s extrinsics are fixed by the body frame of the vehicle, v2s calibration is susceptible to alteration due to the fact that the VGCS's XZ plane remains parallel to the ground surface plane, even when the vehicle's body frame is tilted due to an imbalance of loads.

4.2. Convergence Time

To evaluate the repeatability and consistency of our method, we conducted tests on 60 driving sequences, each lasting five minutes, recorded under identical conditions using the same vehicle. We computed the converged Euler angles from these sequences. We plotted the percentage of converged sequences for thresholds of 0.3, 0.5, 0.7, and 1.0 degrees, as shown in Fig. 6. For a 0.5-degree threshold, roll convergence appears to plateau at 5000 frames, whereas

Date	Sequence	Frames	Pitch						Yaw						Roll						
			E3DC			SX2V			E3DC			SX2V			E3DC			SX2V			
			cam2	cam3	$ \Delta $	cam2	cam3	$ \Delta $	cam2	cam3	$ \Delta $	cam2	cam3	$ \Delta $	cam2	cam3	$ \Delta $	cam2	cam3	$ \Delta $	
10-03	00	4541	0.626	0.596	0.031	-0.905	-0.849	0.057	-0.163	-0.192	0.029	-1.095	-1.299	0.204	0.015	-1.042	1.056	2.254	2.054	0.200	
	01	1099	0.678	0.597	0.081	-0.775	-0.785	0.010	-0.197	-0.221	0.024	-0.814	-0.915	0.101	-1.538	-1.865	0.327	-0.846	-0.910	0.064	
	02	4659	0.636	0.646	0.011	-0.928	-0.873	0.055	-0.267	-0.180	0.087	1.683	1.433	0.250	-1.382	-1.151	0.230	1.947	2.260	0.313	
	avg. std. dev.	-	0.647	0.613	0.034	-0.869	-0.836	0.034	-0.209	-0.198	0.011	-0.075	-0.260	0.185	-0.968	-1.353	0.384	1.119	1.135	0.016	
09-26	03	799	-0.220	-0.188	0.032	-0.305	1.959	2.264	-0.354	-0.214	0.140	-0.780	-5.728	4.948	0.201	0.709	0.508	-0.637	-0.133	0.504	
09-30	04	269	0.905	0.888	0.017	-1.311	-1.271	0.040	-0.384	-0.311	0.073	-0.666	-0.851	0.186	1.342	-1.270	2.612	1.404	1.586	0.182	
	05	2759	0.937	0.955	0.018	-1.017	-1.038	0.021	-0.250	-0.355	0.105	1.250	1.044	0.206	-1.153	-0.948	0.206	1.650	1.631	0.018	
	06	1099	0.966	0.984	0.018	-1.203	-1.191	0.012	-0.216	-0.183	0.033	-2.073	-2.673	0.600	-1.815	-2.137	0.322	1.041	1.886	0.845	
	07	1099	0.940	0.939	0.001	-1.269	-1.229	0.040	-0.332	-0.143	0.189	3.670	3.446	0.224	-0.030	-0.096	0.067	0.396	0.257	0.139	
	08	4069	0.982	0.974	0.008	-1.206	-1.199	0.007	-0.333	-0.293	0.041	0.969	0.809	0.160	-0.317	-0.518	0.201	0.969	1.149	0.180	
	09	1589	0.969	0.941	0.028	-1.256	-1.222	0.034	-0.359	-0.237	0.122	-7.529	-9.806	2.276	1.466	0.891	0.575	0.976	0.847	0.130	
	10	1199	0.850	0.815	0.036	-0.177	-0.784	0.608	-0.068	-0.165	0.096	-1.358	-1.139	0.219	-1.345	-0.634	0.711	0.029	0.031	0.002	
	11	919	0.779	0.737	0.043	-1.198	-1.234	0.036	-0.295	-0.283	0.013	0.325	0.004	0.321	-0.421	-0.700	0.279	0.206	0.799	0.593	
	12	1059	1.046	1.037	0.009	-1.368	-1.244	0.125	-0.126	-0.152	0.025	13.708	-0.378	14.087	-2.568	-2.422	0.146	0.785	1.112	0.327	
	avg. std. dev.	-	0.930	0.919	0.020	-1.112	-1.157	0.103	-0.263	-0.236	0.078	0.922	-1.061	2.031	-0.538	-0.870	0.569	0.828	1.033	0.268	
				0.078	0.092	-	0.364	0.155	-	0.108	0.078	-	5.697	3.691	-	1.354	1.005	-	0.537	0.623	-

Table 2. The results of v2c via c2c on the KITTI dataset for SX2V and E3DC. The bold denotes better calibration result with small absolute error.

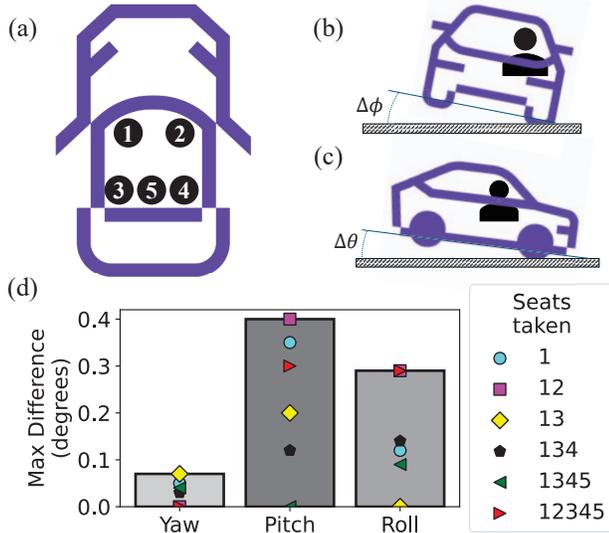


Figure 5. Fluctuation of angles between the VGCS and the SCS due to passengers. (a) Top view of the car. Seat 1: Driver, Seat 2: Passenger, Seat 3: Behind driver, Seat 4: Behind passenger, Seat 5: Back center. (b) Roll angle altered by passenger’s weight. (c) Pitch angle altered by passenger’s weight. (d) Euler angle differences in degrees with different seat configurations. The sequential numbers indicate the seats that were occupied.

pitch and yaw converge around 500 and 1000 frames, respectively. The fast convergence is the result of a combination of two factors. One factor is that the geodesic loss provides a metric for assessing whether an obtained epipole corresponds to the LVP or not. The other factor is that vehicle trajectories are predominantly straight, leading to a probability density function (PDF) of epipoles that peaks near the LVP as demonstrated in Fig. 4. This results in dense sampling near the LVP and faster convergence. Un-

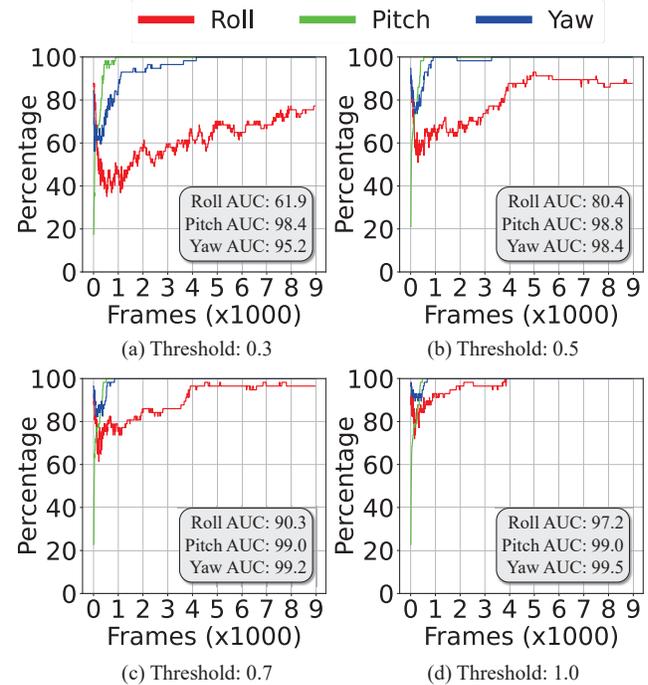


Figure 6. Convergence of roll, pitch, and yaw for a single front facing camera. From 60 sequences, each lasting five minutes, we calculate the final converged Euler angles. With a convergence threshold of (a) 0.3, (b) 0.5, (c) 0.7, and (d) 1.0 degrees, we plot the percentage of sequences within the threshold against the frames progressed.

fortunately, HL estimation is relatively slower since turns occur less frequently, given the same straight trajectories.

Nonetheless, E3DC shows potential for real-world applicability. Tesla’s Autopilot functionality requires “highly-visible lane markings in both the driving lane and adjacent lanes (at least two lanes over on each side of the vehi-

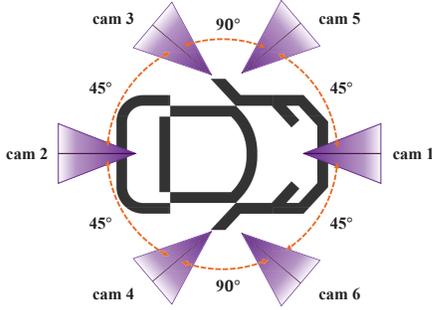


Figure 7. Camera Setup of Custom Dataset

cle)” which takes “2-3 hours [29].” Additionally, NVIDIA’s Driveworks SDK [20] requires an Inertial Measurement Unit (IMU) to measure roll and requires 3 minutes and 30 seconds for pitch and yaw. Our result in Fig. 6 shows pitch and yaw convergence by about 33 seconds (1000 frames/30 fps) and roll convergence by about 2 minutes and 46 seconds (5000 frames/30 fps), 45 seconds less than NVIDIA’s.

4.3. Evaluation Method

To estimate the R_{ji}^c and t_{ji}^c for cameras, we use a simple Visual Odometry (VO) implementation that finds features with Good Features to Track [24] and track them with DIS optical flow [11]. The relative pose is estimated with the 5-point algorithm [18]. We opted for this basic VO methodology to ensure that the computed R_{ji}^c and t_{ji}^c accuracies are representative of the most rudimentary VO pipeline.

Studies focusing on v2s calibration are comparatively limited, with only one related work, *i.e.*, SX2V [36], having open-sourced their approach. SX2V employs supervised deep learning to determine the VPs/HL from environmental cues. However, it only provides code to detect VPs/HL in an image, *i.e.*, a world-to-camera calibration task. Therefore, to determine vehicle-to-world alignment, which in turn results in v2s calibration, we implemented the rest of the algorithm which calculates the Euler angles when the estimated angles in a moving window have a standard deviation below 0.02 radians for a 60-frame time window, as describe in their paper.

KITTI. The KITTI dataset [5] is renowned for its precise s2s calibration. However, it does not provide v2s transformations, preventing direct evaluations for E3DC. Therefore, we indirectly evaluate the accuracy of the v2s calibration based on the s2s calibration results.

Custom Dataset. Extrinsic rotation calibrations are susceptible to changes depending on factors such as tire pressure and the number of passengers, as shown in Fig. 5. This inevitable uncertainty in currently available datasets prompted us to create our own accurate dataset. We employ Perspective-n-Point [4] to get the GT. We use laser line pointers to align an edge of a checkerboard with the vehicle’s rear axle. With measured XYZ coordinates of the camera and checkerboard relative to the vehicle and the 2D

annotated corner points in the image, we optimize for the extrinsic rotation of the camera. We repeat this 5 times to ensure an accurate measure. To ensure that the extrinsic rotations are unaffected by passenger loads, the calibration procedure and recording were performed with the same driver seated in the vehicle.

As depicted in Fig. 7, the vehicle is equipped with three cameras towards the front at 45-degree intervals with cam 1 facing the front and three cameras towards the back at 45-degree intervals with cam 2 facing the back. A The dataset encompasses two distinct environments: an urban neighborhood and a parking lot, with each location providing about 9000 frames from a five-minute drive.

4.4. Quantitative Results

Results on KITTI. Based on the analysis in Fig. 5, we grouped sequences by date. While yaw and roll estimates may not have fully converged due to the limited number of long sequences, pitch estimates are highly accurate and converge quickly. Although KITTI does not provide groundtruth v2s calibrations, we observed the consistency of pitch estimates, where the standard deviation of pitch is merely 0.028, 0.029, 0.078, and 0.092 for cam2 and cam3 in sequences 10-03 and 09-30. The results on KITTI are shown in Tab. 2. Notably, the average pitch estimates on 10-03 is 0.647 and 0.613, while they are 0.930 and 0.919 on 09-30, differing for each day, which reinforces the claim that v2s calibration is susceptible to subtle but meaningful changes. E3DC also shows low absolute yaw

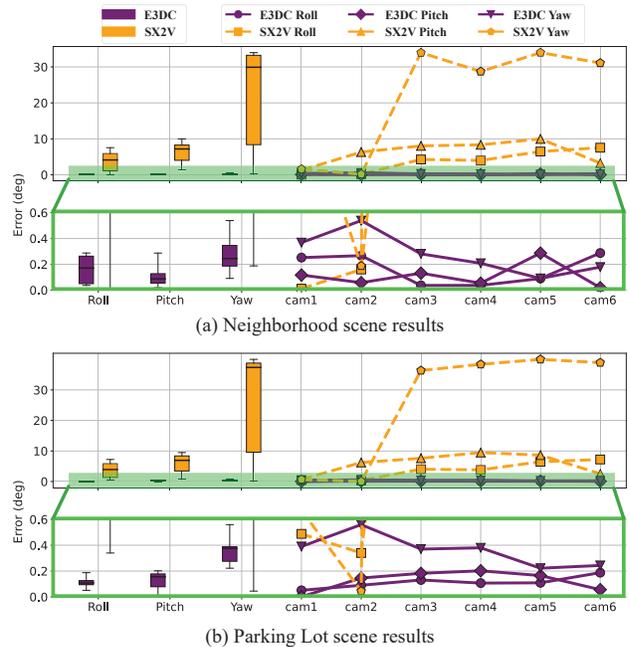


Figure 8. Results from two scenes: Neighborhood and Parking Lot. The green box indicates the zoomed-in region.

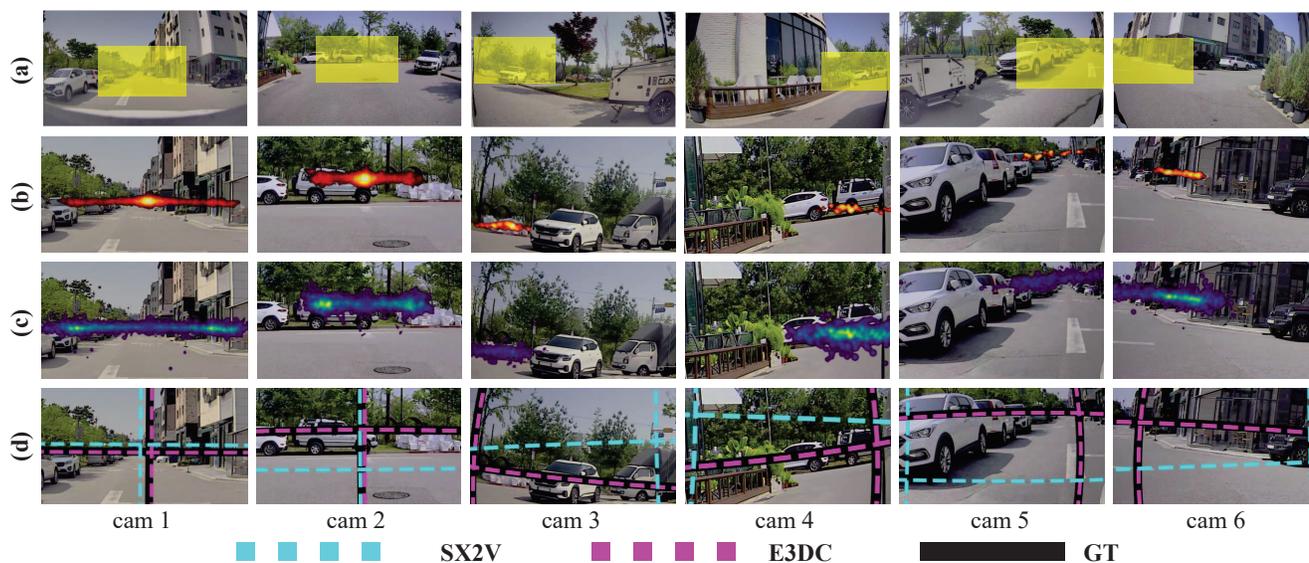


Figure 9. **Visual results on our custom dataset.** (a) Images from the corresponding cameras. The yellow box indicates the zoomed-in region for (b-d). (b) Heat map of the epipoles found throughout a video sequence. (c) Geodesic loss of the epipoles. (d) The horizontal line represents the HL, and the intersecting point represents the LVP.

differences between the two cameras, with the highest being 0.140 degrees for sequence 03. The absolute difference of 0.140 degrees could result from premature yaw estimates, as sequence 03 only has 799 frames, slightly short of the 1000 frames required for yaw convergence. Nonetheless, these results are far better than SX2V’s yaw estimates, where absolute yaw differences reach 4.948 degrees in sequence 03, and the standard deviation of yaw estimates from the same date from 09–30 reaches 5.697, reflecting the highly inconsistent and inaccurate results. E3DC’s roll performance, on the other hand, is relatively lacking, exhibiting higher standard deviations in sequences from 09–30. However, it is important to note that the majority of sequences (03, 04, 06, 07, 10, 11, 12) comprise approximately 1000 frames or fewer. As indicated in Fig. 6, this duration is insufficient for E3DC’s roll estimate to converge; convergence is typically observed at around 4500 frames, corresponding to the longest sequence in the dataset.

Results on Custom Dataset. Fig. 8 shows that E3DC outperforms SX2V in all viewing directions with median roll, pitch, yaw errors of 0.17, 0.09, and 0.24 degrees, respectively, for the *Neighborhood* scene. On the other hand, SX2V initially failed to find a proper yaw candidate due to the frequent turns made by the vehicle, highlighting a key weakness of this method: SX2V requires the vehicle to traverse a straight road with sufficient lines for a long enough duration to produce accurate yaw estimates. Consequently, we gradually increased the standard deviation threshold from 0.02 radians in steps of 0.02 radians until a yaw estimate was retrieved. The resulting median roll, pitch, yaw errors of SX2V in the *Neighborhood* scene are 4.13, 7.19, and 29.91 degrees, respectively.

SX2V was trained on the front view of a camera. We were curious about how transferable SX2V is to different viewing directions such as camera 3,4,5, and 6. Roll estimations for camera 1 and camera 2 seem to be reasonable with only 0.01 and 0.16 degree errors in the *Neighborhood* scene. However, 45-degree angle views for cameras 3,4,5, and 6 resulted in roll estimation errors of up to 7.54 degrees. While pitch estimation also suffered, with an average error of 7.40 degrees, yaw estimations were the most affected, with errors of 33.95, 28.72, 33.97, and 31.09 degrees for cameras 3,4,5, and 6, respectively. This demonstrates that deep learning networks lack transferability across different viewing directions and need to be trained on all views. In contrast, E3DC shows low average errors across all views. The qualitative results of E3DC on the custom dataset are showcased in Fig. 9.

5. Conclusion

This paper introduces E3DC, a sensor-agnostic v2s calibration method, and demonstrates state-of-the-art performance on two datasets. A key attribute of E3DC is its remarkable simplicity, requiring only the attachment of sensors to the vehicle and standard driving procedures for successful calibration. We offer the first v2s calibration dataset to highlight the importance of v2s calibration.

6. Acknowledgements

This research was supported by StradVision. We appreciate all the supports of StradVision members who provided insight and expertise. The contents are solely the responsibility of the authors.

References

- [1] Martin Armstrong, Andrew Zisserman, and Richard Hartley. Self-calibration from image triplets. In *Eur. Conf. Comput. Vis.*, pages 1–16, 1996. 4
- [2] Paul A. Beardsley and Andrew Zisserman. Affine calibration of mobile vehicles. In *Europe-China Workshop on Geometrical Modelling and Invariants for Computer Vision*, pages 214–221, 1995. 4
- [3] Zaipeng Duan, Xuzhong Hu, Junfeng Ding, Pei An, Xiao Huang, and Jie Ma. A robust lidar-camera self-calibration via rotation-based alignment and multi-level cost volume. *IEEE Robot. Autom. Lett.*, 9(1):627–634, 2024. 1
- [4] Xiao-Shan Gao, Xiao-Rong Hou, Jianliang Tang, and Hang-Fei Cheng. Complete solution classification for the perspective-three-point problem. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(8):930–943, 2003. 7
- [5] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The KITTI dataset. *Int. J. Robot. Res.*, 32(11):1231–1237, 2013. 1, 2, 7
- [6] Jakob Geyer, Yohannes Kassahun, Mentar Mahmudi, Xavier Ricou, Rupesh Durgesh, Andrew S. Chung, Lorenz Hauswald, Viet Hoang Pham, Maximilian Mühlegg, Sebastian Dorn, Tiffany Fernandez, Martin Jänicke, Sudesh Mirashi, Chiragkumar Savani, Martin Sturm, Oleksandr Vorobiov, Martin Oelker, Sebastian Garreis, and Peter Schubert. A2D2: Audi Autonomous Driving Dataset, 2020. 1
- [7] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 4
- [8] Jinbeum Jang, Youngran Jo, Minwoo Shin, and Joonki Paik. Camera orientation estimation using motion-based vanishing point detection for advanced driver-assistance systems. *IEEE Trans. Intell. Transp. Syst.*, 22(10):6286–6296, 2021. 1, 2
- [9] Jinyong Jeong, Younghun Cho, and Ayoung Kim. The road is enough! extrinsic calibration of non-overlapping stereo camera and lidar using road information. *IEEE Robot. Autom. Lett.*, 4(3):2831–2838, 2019. 1
- [10] Yonglin Jing, Chongjian Yuan, and Xiaoping Hong. Online calibration between camera and lidar with spatial-temporal photometric consistency. *IEEE Robot. Autom. Lett.*, 9(2):1027–1034, 2024. 1
- [11] Till Kroeger, Radu Timofte, Dengxin Dai, and Luc Van Gool. Fast optical flow using dense inverse search. In *Eur. Conf. Comput. Vis.*, pages 471–488, 2016. 7
- [12] Jung Hyun Lee and Dong-Wook Lee. A hough-space-based automatic online calibration method for a side-rear-view monitoring system. *Sensors*, 20(12), 2020. 1, 2
- [13] Jeong-Kyun Lee, Young-Ki Baik, Hankyu Cho, and Seungwoo Yoo. Online extrinsic camera calibration for temporally consistent ipm using lane boundary observations with a lane width prior, 2020. 1, 2
- [14] Binbin Li, Xinyu Du, Yao Hu, Hao Yu, and Wende Zhang. Online camera-to-ground calibration for autonomous driving, 2023. 1, 2
- [15] Xiyuan Liu and Fu Zhang. Extrinsic calibration of multiple lidars of small fov in targetless environments. *IEEE Robot. Autom. Lett.*, 6(2):2036–2043, 2021. 1
- [16] Michael Miksch, Bin Yang, and Klaus Zimmermann. Automatic extrinsic camera self-calibration based on homography and epipolar geometry. In *IEEE Intell. Vehicles Symp.*, pages 832–839, 2010. 1, 2
- [17] Yue Ming, Xuyang Meng, Chunxiao Fan, and Hui Yu. Deep learning for monocular depth estimation: A review. *Neurocomputing*, 438:14–33, 2021. 1
- [18] D. Nister. An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(6):756–770, 2004. 7
- [19] Michał R. Nowicki. Spatiotemporal calibration of camera and 3d laser scanner. *IEEE Robot. Autom. Lett.*, 5(4):6451–6458, 2020. 1
- [20] NVIDIA. Feature-based camera self-calibration. https://docs.nvidia.com/drive/driveworks-4.0/calibration_usecase_features.html. [Accessed 30-01-2024]. 7
- [21] Zhanpeng Ouyang, Lan Hu, Yukan Lu, Zhirui Wang, Xin Peng, and Laurent Kneip. Online calibration of exterior orientations of a vehicle-mounted surround-view camera system. In *IEEE Int. Conf. Robot. Autom.*, pages 4990–4996, 2020. 2
- [22] Chanoh Park, Peyman Moghadam, Soohwan Kim, Sridha Sridharan, and Clinton Fookes. Spatiotemporal camera-lidar calibration: A targetless and structureless approach. *IEEE Robot. Autom. Lett.*, 5(2):1556–1563, 2020. 1
- [23] Long Sha, Jennifer Hobbs, Panna Felsen, Xinyu Wei, Patrick Lucey, and Sujoy Ganguly. End-to-end camera calibration for broadcast videos. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 13624–13633, 2020. 1
- [24] Jianbo Shi and Tomasi. Good features to track. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 593–600, 1994. 7
- [25] Jieying Shi, Ziheng Zhu, Jianhua Zhang, Ruyi Liu, Zhenhua Wang, Shengyong Chen, and Honghai Liu. Calibrcnn: Calibrating camera and lidar by recurrent convolutional neural network and geometric constraints. In *IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pages 10197–10202, 2020. 1
- [26] Wonho Song, Minh Oh, Jaeyoung Lee, and Hyun Myung. Galibr: Targetless lidar-camera extrinsic calibration method via ground plane initialization. In *IEEE Intell. Vehicles Symp.*, pages 217–223, 2024. 1
- [27] Junzhe Su, Masahiro Hirano, and Yuji Yamakawa. Online camera orientation calibration aided by a high-speed ground-view camera. *IEEE Robot. Autom. Lett.*, 8(10):6275–6282, 2023. 1, 2
- [28] Jun Tan, Xiangjing An, Xin Xu, and Hangen He. Automatic extrinsic calibration for an onboard camera. In *Chin. Autom. Congr.*, pages 340–343, 11 2013. 1, 2
- [29] TESLA. Calibrating cameras. https://service.tesla.com/docs/Public/diy/modely/en_us/GUID-D6F7D1BC-193D-4A2E-99B0-E3BA3D41BDE2.html. [Accessed 30-01-2024]. 7
- [30] Ivana Tošić and Pascal Frossard. Chapter 10 - spherical imaging in omnidirectional camera networks. In *Multi-Camera Networks*, pages 239–264. Academic Press, 2009. 4

- [31] Gopi Krishna Tummala, Tanmoy Das, Prasun Sinha, and Rajiv Ramnath. Smartdashcam: Automatic live calibration for dashcams. In *ACM/IEEE Int. Conf. Inf. Process. Sensor Netw.*, pages 157–168, 2019. [1](#), [2](#)
- [32] Rafael Grompone Von Gioi, Jérémie Jakubowicz, Jean-Michel Morel, and Gregory Randall. Lsd: A line segment detector. *Image Processing On Line*, 2:35–55, 2012. [2](#)
- [33] Jens Westerhoff, Stephanie Lessmann, Mirko Meuter, and Anton Kummert. A classification and temporal filtering based system for online extrinsic camera calibration. In *IEEE Int. Conf. Intell. Transp. Syst.*, pages 1285–1290, 2016. [1](#), [2](#)
- [34] Wanting Xu, Si’Ao Zhang, Li Cui, Xin Peng, and Laurent Kneip. Event-based visual odometry on non-holonomic ground vehicles. In *Int. Conf. 3DV*, pages 831–841, 2024. [4](#)
- [35] Guohang Yan, Feiyu He, Chunlei Shi, Pengjin Wei, Xinyu Cai, and Yikang Li. Joint camera intrinsic and lidar-camera extrinsic calibration. In *IEEE Int. Conf. Robot. Autom.*, pages 11446–11452, 2023. [1](#)
- [36] Guohang Yan, Zhaotong Luo, Zhuochun Liu, Yikang Li, Botian Shi, and Kai Zhang. Sensorx2vehicle: Online sensors-to-vehicle rotation calibration methods in road scenarios. *IEEE Robot. Autom. Lett.*, 9(4):3775–3782, 2024. [1](#), [2](#), [7](#)
- [37] Chongjian Yuan, Xiyuan Liu, Xiaoping Hong, and Fu Zhang. Pixel-level extrinsic self calibration of high resolution lidar and camera in targetless environments. *IEEE Robot. Autom. Lett.*, 6(4):7517–7524, 2021. [1](#)
- [38] Guanyu Zhang, Kunyang Wu, Jun Lin, Tianhao Wang, and Yang Liu. Automatic extrinsic parameter calibration for camera-lidar fusion using spherical target. *IEEE Robot. Autom. Lett.*, 9(6):5743–5750, 2024. [1](#)
- [39] Xinyu Zhang, Shifan Zhu, Shichun Guo, Jun Li, and Huaping Liu. Line-based automatic extrinsic calibration of lidar and camera. In *IEEE Int. Conf. Robot. Autom.*, pages 9347–9353, 2021. [1](#)
- [40] Youcheng Zhang, Zongqing Lu, Xuechen Zhang, Jing-Hao Xue, and Qingmin Liao. Deep learning in lane marking detection: A survey. *IEEE Trans. Intell. Transp. Syst.*, 23(7):5976–5992, 2022. [1](#)
- [41] Lipu Zhou, Zimo Li, and Michael Kaess. Automatic extrinsic calibration of a camera and a 3d lidar using line and plane correspondences. In *IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pages 5562–5569, 2018. [1](#)
- [42] Yuewen Zhu, Chunran Zheng, Chongjian Yuan, Xu Huang, and Xiaoping Hong. Camvox: A low-cost and accurate lidar-assisted visual slam system. In *IEEE Int. Conf. Robot. Autom.*, pages 5049–5055, 2021. [1](#)
- [43] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo, and Jieping Ye. Object detection in 20 years: A survey. *Proceedings of the IEEE*, 111(3):257–276, 2023. [1](#)
- [44] David Zuñiga-Noël, Jose-Raul Ruiz-Sarmiento, Ruben Gomez-Ojeda, and Javier Gonzalez-Jimenez. Automatic multi-sensor extrinsic calibration for mobile robots. *IEEE Robot. Autom. Lett.*, 4(3):2862–2869, 2019. [1](#)