TEScalib: Targetless Extrinsic Self-Calibration of LiDAR and Stereo Camera for Automated Driving Vehicles with Uncertainty Analysis

Haohao Hu[*,1], Fengze Han[1], Frank Bieder[2], Jan-Hendrik Pauls[1] and Christoph Stiller[1,2]

Abstract— In this paper, we present TEScalib, a novel extrinsic self-calibration approach of LiDAR and stereo camera using the geometric and photometric information of surrounding environments without any calibration targets for automated driving vehicles. Since LiDAR and stereo camera are widely used for sensor data fusion on automated driving vehicles, their extrinsic calibration is highly important. However, most of the LiDAR and stereo camera calibration approaches are mainly target-based and therefore time consuming. Even the newly developed targetless approaches in last years are either inaccurate or unsuitable for driving platforms.

To address those problems, we introduce TEScalib. By applying a 3D mesh reconstruction-based point cloud registration, the geometric information is used to estimate the LiDAR to stereo camera extrinsic parameters accurately and robustly. To calibrate the stereo camera, a photometric error function is builded and the LiDAR depth is involved to transform key points from one camera to another. During driving, these two parts are processed iteratively. Besides that, we also propose an uncertainty analysis for reflecting the reliability of the estimated extrinsic parameters. Our TEScalib approach evaluated on the KITTI dataset achieves very promising results.
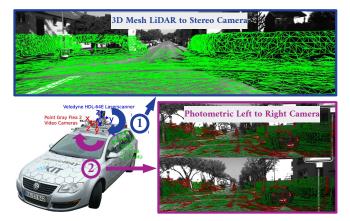
Fig. 1: TEScalib: a targetless LiDAR stereo camera extrinsic calibration approach for automated driving platforms. By applying a 3D mesh reconstruction-based point cloud registration and a photometric error function, the LiDAR and stereo camera are calibrated accurately and robustly.

I. INTRODUCTION

Autonomous systems like robots and automated driving vehicles are often equipped with multi-modal sensors for environment perception. A sensor setup composed of different sensor measuring principles make perception robust to the variety of environments. For fusing measurements of different sensor systems, it is a key requirement to firstly obtain the extrinsic parameters between them accurately. Most common extrinsic calibration approaches for LiDAR and stereo camera use calibration targets such as chessboards or other objects with a known shape. The extrinsic parameters are estimated by minimizing the back-projection error of the corresponding calibration targets in the LiDAR or camera domain. An experimental environment with enough calibration targets is usually needed, which introduces much manual work. However, this manual work can not be easily done by customers themselves, because some professional knowledge is also needed. Especially for custom automated driving vehicles in the future, the mounted sensors need to be re-calibrated continuously after they are sold.

In the work [1], the authors propose a targetless LiDAR and camera extrinsic calibration approach based on mutual information (MI) using LiDAR intensity measurements and camera gray values. Cameras measure rich photometric information and LiDARs accurate geometric information of the surroundings. Only using the inaccurate LiDAR intensity to obtain the extrinsic parameters between LiDAR and camera seems unpromising. The work in [2] co-registers the camera edge pixels and LiDAR discontinuity features. However, it faces a parallax issue since sensors can not be mounted with a same pose. The co-registered features do not always represent a same feature in the real world. Besides that, all targetless approaches just give us a calibration result in the end, but can not decide whether the result is reliable.

Therefore, as shown in Fig. 1 we propose TEScalib: a novel targetless extrinsic self-calibration approach of LiDAR and stereo camera using both geometric and photometric information of the surrounding environments. Our TEScalib contains mainly four parts: LiDAR-stereo calibration, left and right camera calibration, cooperative calibration (so called co-calibration) and uncertainty analysis. As shown in Fig. 2, for the LiDAR-stereo calibration, a 3D mesh reconstruction-based robust point cloud registration method is applied, which achieves a very large convergence range and a high accuracy. A photometric error function using camera image key points and LiDAR depth is minimized to obtain the left and right camera calibration precisely. During driving, both approaches are processed iteratively and an uncertainty analysis is involved to evaluate the reliability of the estimated parameters under different driving scenarios.

[*]Corresponding author.

[1] Institute of Measurement and Control, Karlsruhe Institute of Technology, Karlsruhe, Germany. {haohao.hu, janhendrik.pauls, stiller}@kit.edu, skyhfz@gmail.com

[2] Mobile Perception Systems Department, FZI Research Center for Information Technology, Karlsruhe, Germany. bieder@fzi.de
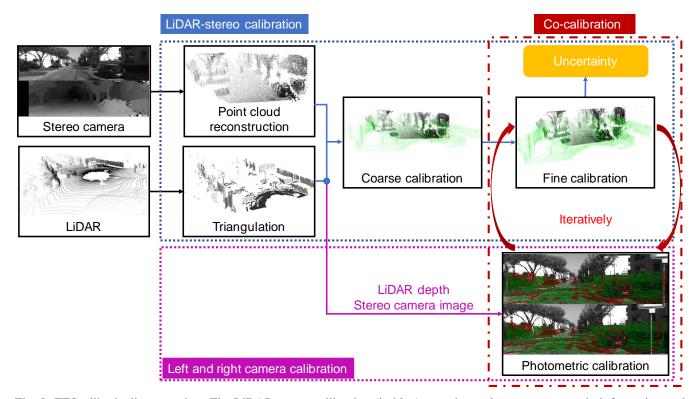
Fig. 2: TEScalib pipeline overview. The LiDAR-stereo calibration (in blue) uses the environment geometric information and achieves robust and accurate results by applying a 3D mesh reconstruction-based point cloud registration method. By the left and right camera calibration (in violet), camera key points and LiDAR depth are used to build a photometric error function, by minimizing which the transformation between two cameras is estimated accurately. These two parts run cooperatively and iteratively (in red). An uncertainty analysis (in orange) is done to obtain the reliability of the estimated extrinsic parameters.

The main contributions of this work can be listed as:

- LiDAR-stereo calibration using 3D mesh reconstruction achieves a large convergence range and a high accuracy.
- Left and right camera calibration with a photometric error function is sensitive to camera poses but not to the key point depth from the LiDAR 3D meshes, which makes an online iterative co-calibration beginning with a bad LiDAR depth initialization meaningful.
- An uncertainty analysis helps to evaluate the reliability of the estimated extrinsic parameters in each single dimension separately under different driving scenarios.

II. RELATED WORK

The targetless extrinsic calibration of the LiDAR and camera sensor is the problem to estimate the rigid body transformation between the LiDAR and camera coordinate systems without any calibration targets. The existing approaches can be mainly divided into 2 categories: the motion-based approaches and the feature-based approaches.

For the motion-based approaches like [3], [4], an initialization is not required to ensure the convergence. Chanoh et al. [5] use the motion estimated from a LiDAR odometry to transform camera feature points into a single coordinate system in order to build a back-projection error function, by minimizing which the extrinsic parameters between the

LiDAR and camera can be obtained. However the motion-based approaches achieve usually low accuracy because of the existing noise inside of the estimated motion. Besides that, in order to fix the motion degeneration problem, it requires at least 2 Degrees of Freedom (DoF) rotational motion to make all translational extrinsic parameters observable.

The feature-based approaches need an initialization and can usually achieve a higher accuracy. In [1], [6], the extrinsic parameters between the LiDAR and camera are estimated by maximizing the MI obtained from the sensor-measured surface intensities. However the influence of intensities by sunlight or shadow can damage the calibration results. The works proposed in [7], [8] estimate extrinsic parameters by aligning edges or lines detected in both LiDAR and camera domains. In [9], the LiDAR Simultaneous Localization and Mapping (SLAM) technique is used to map the road surface. The mapped road surface features are then back-projected into images in order to build residuals. The authors in [10] apply a motion-based approach to estimate the extrinsic parameters coarsely and use the feature-based method to refine the result. However finding the right feature correspondences from different sensors is a big challenge and wrong associations can strongly reduce the estimation accuracy. Besides that, the detected features like lines from the LiDAR and camera do not always represent the same

feature in the real world due to the parallax effect.

Another big challenge of a targetless LiDAR and camera extrinsic calibration algorithm is the reliability analysis of the estimated calibration result. Due to the lack of the constraints generated from different driving scenarios, residuals after optimization could be quite small but the elements in some specified dimensions of the calibration result could still have large errors. The recent targetless calibration approaches lack a mechanism to make an uncertainty analysis of the estimated extrinsic parameters, which could be very helpful to evaluate the reliability of the calibration result on the one hand. An the other hand, it is also very meaningful for an automatic selection of driving scenarios for an optimal calibration with small uncertainty in all 6 DoF.

To address the problems, we propose TEScalib. For the LiDAR-stereo calibration, instead of only using special features like lines or edges, we use all geometric information of driving surroundings. No explicit association is needed and thanks to the LiDAR 3D mesh reconstruction we achieve a very large convergence and a high accuracy. For the left and right camera calibration, the accurate LiDAR depth reduces the false key point association ratio, which makes the calibration robust and accurate. Besides that, the uncertainty analysis evaluates the reliability of the estimated extrinsic parameters under different driving scenarios.

III. LiDAR-stereo Calibration

The LiDAR-stereo calibration contains three steps: data preprocessing, coarse calibration and fine calibration. By the data preprocessing step, the 3D environment geometry is extracted from the LiDAR and stereo camera measurements. The goal of the coarse calibration is to estimate the extrinsic parameters with a large convergence range robustly based on a bad initialization. Afterwards the fine calibration refines the extrinsic parameters with a very high accuracy.

A. Data Preprocessing

To extract the 3D environment geometry from stereo camera images, a dense disparity image is calculated using the Semi-Global Block Matching (SGBM) approach presented in [11] implemented in *OpenCV* [12]. Subsequently, we use the known camera intrinsic parameters to reconstruct the stereo camera 3D point clouds from these estimated dense disparity images. Instead of using points directly, 3D reconstructed triangle meshes are used to represent environment geometric primitives from the LiDAR measurements, which helps us to stabilize our data association process during the coarse calibration and increase the convergence range.

B. Coarse Calibration

Generally, the coarse calibration is responsible for estimating extrinsic parameters with a high convergence range and a high robustness, which is however strongly effected by the correct rate of data association. Because of mismatches and holes in disparity images and the large initial bias of extrinsic parameters, which lead to an inaccurate association, a robust point-to-plane data association strategy is developed.
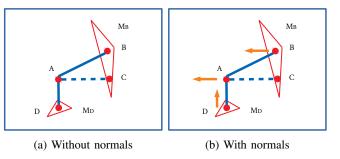

(a) Without normals   (b) With normals

Fig. 3: Strategies to calculate the point triangle face distance.

The traditional Iterative Closest Point (ICP) algorithm only uses the Euclidean distance to search the nearest neighbors and ignores the surface normal information. Since there are many large planes in urban environments, the surface normals can provide extra information to increase the data association quality. Therefore we add the surface normal information into our association searching process and the adapted searching criterion is defined as:

$$\arg\min_{^{C}\mathbf{p}_i \in \mathcal{N}_c} \|(^{C}\mathbf{p}_i - ^{L}\mathbf{p}_i) + \omega \cdot (^{C}\mathbf{n}_i - ^{L}\mathbf{n}_i)\|^2, \quad (1)$$

where $\mathcal{N}_c$ denotes the stereo point cloud, $^{C}\mathbf{p}_i$ denotes a stereo point and $^{L}\mathbf{p}_i$ the centroid of the corresponding LiDAR triangle face. $^{C}\mathbf{n}_i$ is the normal vector at the stereo point and $^{L}\mathbf{n}_i$ the normal vector of the corresponding LiDAR triangle mesh. ω is a scalar hyperparameter to balance the importance of the center point position and the normal vector. As shown in Fig. 3, the surface normal information provides extra geometric information to achieve better data association results. In Fig. 3a, the stereo point A and centroid D of the LiDAR triangle face M_D have the shortest distance, if only the perpendicular distance from a point to a triangle mesh is considered. The stereo point A is then associated to the LiDAR triangle face M_D. However, if the surface normal is also considered as shown in Fig. 3b, the stereo point A will be associated to the LiDAR triangle face M_B correctly. Therefore, during the point cloud registration, we consider the surface normal information in the objective function:

$$\arg\min_{\mathbf{T}} \sum_{\mathbf{p}_s^i, \mathbf{p}_m^i} \frac{\mathbf{n}_m^i(^{L}_{C}\mathbf{T}\mathbf{p}_m^i - \mathbf{p}_s^i)}{\|\mathbf{n}_m^i\|} + \omega \cdot ^{L}_{C}\mathbf{T}\mathbf{n}_m^i \times \mathbf{n}_s^i, \quad (2)$$

where $\mathbf{n}_s^i$ and $\mathbf{n}_m^i$ denote the normal vector of a stereo point and the normal vector of the corresponding LiDAR triangle mesh. $^{L}_{C}\mathbf{T}$ denotes the LiDAR-stereo extrinsic parameters. $\mathbf{p}_s^i$ is a stereo point and $\mathbf{p}_m^i$ the centroid point of the corresponding LiDAR triangle mesh. Since the normal vector estimation of the ground and big walls is accurate, this objective function makes our coarse calibration accurate.

C. Fine Calibration

During the fine calibration step, a point to point objective function is developed to refine the calibration results accurately. Because the point cloud from the stereo camera is relatively noisy, the corresponding normal vector calculation

is not accurate enough, a point to point Euclidean distance is utilized to search the nearest neighbors. Due to the high accuracy of the LiDAR depth measuring, the normal information of the LiDAR points are used to set the projection direction and the objective function is defined as:

$$\arg\min_{\mathbf{T}} \sum_{\mathbf{p}_i^i, \mathbf{p}_s^i} {}^{L}\mathbf{n}_i^{\mathsf{T}} [{}_{C}^{L}\mathbf{T}({}^{L}\mathbf{p}_i + {}^{L}\mathbf{w}_i) - \mathbf{F}({}^{C}d_i + {}^{C}w_i)], \quad (3)$$

where $\mathbf{F}(.)$ denotes the disparity to point cloud transform function, ${}_{C}^{L}\mathbf{T}$ denotes the extrinsic parameters, ${}^{L}\mathbf{w}_i$ and ${}^{C}w_i$ denote the noise of LiDAR and stereo point clouds.

IV. LEFT AND RIGHT CAMERA CALIBRATION

An important requirement for the LiDAR-stereo calibration is the well calibrated stereo camera setup. To ensure this, we design a stereo left and right camera extrinsic calibration approach with a camera key point-based photometric objective function using the LiDAR depth information.

A. Key Point Projection Function

A LiDAR triangle mesh in the 3D space can be observed from the left and right camera with the projection function:

$$d_r\mathbf{p}_r = \mathbf{K} \cdot {}_{l}^{r}\mathbf{T}\left(\mathbf{K}^{-1}d_l\mathbf{p}_l\right), \quad (4)$$

where $\mathbf{p}_r = (u_r, v_r, 1)^{\mathsf{T}}$ and $\mathbf{p}_l = (u_l, v_l, 1)^{\mathsf{T}}$ denote the left and right camera image key point. d_r and d_l denote the depth of the points $\mathbf{p}_r$ and $\mathbf{p}_l$. $\mathbf{K}$ denotes the camera intrinsic parameters and ${}_{l}^{r}\mathbf{T}$ the stereo extrinsic parameters. An affine relationship exists between the projected triangle in the left and right image, since $\mathbf{K}$ is not an orthonormal rather than a full range matrix. Using the LiDAR triangle meshes, points on the triangle face can be linear interpolated with:

$$\mathbf{p} = \alpha\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c}$$
$$\alpha + \beta + \gamma = 1, \quad (5)$$

where $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are the triangle mesh vertices. Because Eq.5 describes a proportional relation, we can use α, β, γ to find the projection point in another camera image directly.

B. Photometric Objective Function

Similar as in the Stereo Direct Sparse Odometry (DSO) [13], the photometric objective function is denoted as:

$$\mathbf{p}_l = \begin{bmatrix} \alpha & \beta & \gamma \end{bmatrix} \boxtimes \mathbf{K} \cdot {}_{L}^{C}\mathbf{T} \begin{bmatrix} \mathbf{a} & \mathbf{b} & \mathbf{c} \\ 1 & 1 & 1 \end{bmatrix}$$

$$\mathbf{p}_r = \begin{bmatrix} \alpha & \beta & \gamma \end{bmatrix} \boxtimes \mathbf{K} \cdot {}_{l}^{r}\mathbf{T} \cdot {}_{L}^{C}\mathbf{T} \begin{bmatrix} \mathbf{a} & \mathbf{b} & \mathbf{c} \\ 1 & 1 & 1 \end{bmatrix}$$

$$E_{lr} = \sum_{\mathbf{p}_l \in \mathcal{P}_l} \sum_{\tilde{\mathbf{p}}_l \in \mathcal{N}_{\mathbf{p}_l}} \omega_{\tilde{\mathbf{p}}_l} \|\mathbf{I}_r[\tilde{\mathbf{p}}_r] - b_r - \frac{e^{a_r}}{e^{a_l}}(\mathbf{I}_l[\tilde{\mathbf{p}}_l] - b_l)\|_\gamma$$

$$\omega_{\mathbf{p}} = \frac{c^2}{(c^2 + \|\nabla\mathbf{I}(\mathbf{p})\|_2^2)}, \quad (6)$$

where $\mathbf{a} = (a_x, a_y, a_z)^{\mathsf{T}}$, $\mathbf{b} = (b_x, b_y, b_z)^{\mathsf{T}}$ and $\mathbf{c} = (c_x, c_y, c_z)^{\mathsf{T}}$ are the triangle face vertices, α, β and γ are the triangle face based parameterization of the point $\mathbf{p}_l$, $\boxtimes$ denotes that the last row needs to be normalized. $\mathcal{P}_l$ is the left

image point set. $\mathcal{N}_{\mathbf{p}_l}$ is the used image pattern of the point $\mathbf{p}_l$ to calculate photometric errors. $\tilde{\mathbf{p}}_r$ is the projection point of $\mathbf{p}_l$ in the right image. The brightness correction approach presented in [13] is here applied and the corresponding parameters a_l, b_l, a_r and b_r are jointly estimated during the optimization process. $\|\cdot\|_\gamma$ denotes the robust Huber norm and c is a hyperparameter to set the weight.

C. Sensitivity Analysis

To know whether the photometric objective function is suited for the left and right camera calibration using the LiDAR depth in the camera frame, which could be noisy because of a bad LiDAR-stereo calibration initialization, we need to analyze the sensitivity of the objective function to the camera pose and LiDAR depth. We assume that the depth of the same key point in the left and right camera is the same: $d = d_l = d_r$. The sensitivity depends on Eq. 6 to the camera pose ${}_{l}^{r}\mathbf{T}$ and to the LiDAR depth d are denoted as:

$$\mathbf{J}_{{}_{l}^{r}\mathbf{T}} = \frac{1}{d_r}\mathbf{K}\left[{}_{l}^{r}\mathbf{T}\left(\mathbf{K}^{-1}d_l\mathbf{p}_l\right), \mathbf{I}\right] \in \mathbb{R}^{3\times6}$$

$$\mathbf{J}_{d_r} = \frac{1}{d_r^2}\mathbf{K}\,{}_{l}^{r}\mathbf{R}\mathbf{p}_l \in \mathbb{R}^{3\times1}. \quad (7)$$

By applying camera intrinsics, the rotation and translation Jacobian of the camera pose and key point depth are then:

$$\mathbf{J}_{{}_{l}^{r}\mathbf{R}} = \begin{bmatrix} c_x v & f_x\left(\frac{t_x}{d} - u\right) + c_x & f_x v \\ f_y\left(\frac{t_x}{d} - u\right) - c_y & c_y & f_y \\ v & 1 & 0 \end{bmatrix}$$

$$\mathbf{J}_{{}_{l}^{r}\mathbf{T}} = \begin{bmatrix} \frac{f_x}{d} & 0 & \frac{c_x}{d} \\ 0 & \frac{f_y}{d} & \frac{c_y}{d} \\ 0 & 0 & \frac{d}{d} \end{bmatrix}$$

$$\mathbf{J}_{d_r} = \frac{1}{d_r^2}\begin{bmatrix} f_x u + c_x \\ f_y v + c_y \\ 1 \end{bmatrix}. \quad (8)$$

As shown in Eq. 8, the photometric objective function is sensitive to the camera pose but not to the LiDAR depth. Therefore the photometric objective function can tolerant some LiDAR stereo calibration noise and still realize an accurate left and right camera calibration.

V. CO-CALIBRATION

The LiDAR-stereo calibration is robust and accurate, when the stereo camera setup is well calibrated. The left and right camera calibration is accurate and can tolerant some LiDAR stereo camera extrinsic parameter error. During driving, these two approaches are processed cooperatively and iteratively, which is called as co-calibration. We stop this process until both of the extrinsic parameter estimations are stable.

VI. UNCERTAINTY ANALYSIS

Most of LiDAR and camera calibration approaches can not decide automatically whether the calibration result is reliable when the residuals are small after optimization. We propose a novel uncertainty analysis to evaluate the reliability of the estimated extrinsic parameters under different scenarios.

A. Measurement Noise Model

Measurement noise leads to the uncertainty of the extrinsic parameter estimation. To analyze the estimation uncertainty, sensor noise models are needed. The stereo point cloud is obtained from the disparity image. Let ${}^{C}\mathrm{w}_i \in \mathcal{N}(0, {}^{C}\Sigma_i)$ be the disparity noise of each pixel. The covariance matrix is ${}^{C}\Sigma_i = \sigma_i^2 \mathbf{I}_{1\times 1}$. The disparity of each pixel can be described as ${}^{C}d_i^{\mathrm{gt}} = {}^{C}d_i + {}^{C}\mathrm{w}_i$. Similar as [14], let ${}^{L}\mathbf{w}_i$ be the measurement noise of the LiDAR point ${}^{L}\mathbf{p}_i$. Based on the measuring principle of LiDAR sensors, this noise can be divided into the beam direction noise and the distance noise assuming that the beam direction has a perturbation in the tangent plane. Let ${}^{l}\mathbf{w}_i \in \mathcal{N}(\mathbf{0}, {}^{l}\Sigma_i)$ be the beam direction noise. The covariance matrix is ${}^{l}\Sigma_i = \sigma_i^2 \mathbf{I}_{2\times 2}$. The beam direction $\boldsymbol{\omega}_i^{\mathrm{gt}} = e^{\lfloor N(\boldsymbol{\omega}_i)^l \mathbf{w}_i \times \rfloor} \boldsymbol{\omega}_i$, where $\lfloor \times \rfloor$ denotes the skew-symmetric matrix mapping, $N(\boldsymbol{\omega}_i) = [\mathbf{N}_1, \mathbf{N}_2] \in \mathbb{R}^{3\times 2}$ is the orthonormal basis of the tangent space at $\boldsymbol{\omega}_i$. Let ${}^{d}\mathrm{w}_i \in \mathcal{N}(0, {}^{d}\Sigma_i)$ be the distance noise, where ${}^{d}\Sigma_i \in \mathbb{R}^{1\times 1}$. The distance is described as ${}^{l}d_i^{\mathrm{gt}} = {}^{l}d_i + {}^{d}\mathrm{w}_i$. Combining these two noise sources, the ground-truth LiDAR point ${}^{L}\mathbf{p}_i^{\mathrm{gt}}$ combined with its measurement ${}^{L}\mathbf{p}_i$ can be described as:

$$
\begin{aligned}
{}^{L}\mathbf{p}_{\mathrm{gt}}^{i} &= {}^{l}d_i^{\mathrm{gt}} \cdot \boldsymbol{\omega}_i^{\mathrm{gt}} \\
&\approx \underbrace{{}^{l}d_i \cdot \boldsymbol{\omega}_i}_{{}^{L}\mathbf{p}_i} + \underbrace{{}^{d}\mathrm{w}_i \cdot \boldsymbol{\omega}_i - {}^{l}d_i \lfloor \boldsymbol{\omega}_i \times \rfloor N(\boldsymbol{\omega}_i)^l \mathbf{w}_i}_{{}^{L}\mathbf{w}_i}.
\end{aligned} \quad (9)
$$

Therefore, the covariance matrix of the LiDAR point measurement noise ${}^{L}\mathbf{w}_i \in \mathcal{N}(\mathbf{0}, {}^{L}\Sigma_i)$ can be described as:

$$
\begin{aligned}
{}^{L}\boldsymbol{\Sigma}_i &= \mathbf{J}_i \begin{bmatrix} {}^{d}\Sigma_i & \mathbf{0}_{1\times 2} \\ \mathbf{0}_{2\times 1} & {}^{l}\boldsymbol{\Sigma}_i \end{bmatrix} \mathbf{J}_i^{\mathsf{T}} \\
\mathbf{J}_i &= [\boldsymbol{\omega}_i, \; -{}^{l}d_i \lfloor \boldsymbol{\omega}_i \times \rfloor N(\boldsymbol{\omega}_i)] \in \mathbb{R}^{3\times 3}.
\end{aligned} \quad (10)
$$

B. Uncertainty Analysis under Different Driving Scenarios

Based on the constraints generated from different driving scenarios, the estimated extrinsic parameters can have high accuracy in some dimensions and low accuracy in other dimensions. Therefore, we introduce a novel uncertainty analysis approach to evaluate the reliability of the LiDAR-stereo calibration result. The corresponding objective function Eq.3 can be approximated as:

$$
\begin{aligned}
0 &= {}^{L}\mathbf{n}_i^{\mathsf{T}} [{}^{L}_{C}\mathbf{T}({}^{L}\mathbf{p}_i + {}^{L}\mathbf{w}_i) - \mathbf{F}({}^{C}d_i + {}^{C}w_i)] \\
&\approx \mathbf{res}_i + \mathbf{J}_{\mathbf{T}_i}\delta {}^{L}_{C}\mathbf{T} + \mathbf{J}_{\mathbf{w}_i}\delta \mathbf{w}_i,
\end{aligned} \quad (11)
$$

where

$$
\begin{aligned}
\mathbf{res}_i &= {}^{L}\mathbf{n}_i^{\mathsf{T}} [{}^{L}_{C}\mathbf{T}({}^{L}\mathbf{p}_i) - \mathbf{F}({}^{C}d_i)] \\
{}^{L}_{C}\mathbf{T} &= \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}_{3\times 1} & 0_{1\times 1} \end{bmatrix} \\
\mathbf{J}_{\mathbf{T}_i} &= {}^{L}\mathbf{n}_i^{\mathsf{T}} \left[-\lfloor \mathbf{R}^L \mathbf{p}_i \rfloor, \; \mathbf{I} \right] \\
\mathbf{J}_{\mathbf{w}_i} &= {}^{L}\mathbf{n}_i^{\mathsf{T}} \left[\mathbf{R}, \; -\frac{\partial \mathbf{F}}{\partial d_i} \right] \\
\mathbf{w}_i &= \begin{bmatrix} {}^{L}\mathbf{w}_i \\ {}^{C}\mathrm{w}_i \end{bmatrix} \in \mathbb{R}^{4\times 1} \\
\boldsymbol{\Sigma}_i &= \begin{bmatrix} {}^{L}\boldsymbol{\Sigma}_i & \mathbf{0} \\ \mathbf{0} & {}^{C}\Sigma_i \end{bmatrix} \in \mathbb{R}^{4\times 4}.
\end{aligned} \quad (12)
$$

Assuming no bias in the left and right camera calibration, the uncertainty of the LiDAR-stereo calibration only depends on the noise of LiDAR points and pixel disparities. Based on the forward and backward propagation, the LiDAR-stereo calibration uncertainty can be defined as:

$$
{}^{L}_{C}\boldsymbol{\Sigma}_{\mathbf{T}} = \left[\sum_{{}^{L}\mathbf{p}_i, {}^{C}d_i} \mathbf{J}_{\mathbf{T}_i}^{\mathsf{T}} \left(\mathbf{J}_{\mathbf{w}_i} \boldsymbol{\Sigma}_i \mathbf{J}_{\mathbf{w}_i}^{\mathsf{T}} \right)^{-1} \mathbf{J}_{\mathbf{T}_i} \right]^{-1}. \quad (13)
$$

VII. EXPERIMENTAL EVALUATION
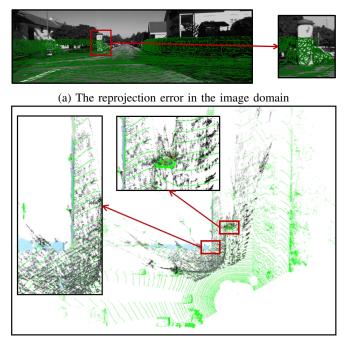
A. Dataset and Parameters

For the experimental evaluation, we use the urban sequence 00 from the KITTI dataset [15]. For comparison, the ground-truth extrinsic parameters between the LiDAR and stereo camera setup as well as between the left and right stereo camera provided by the KITTI dataset are used. We also compare our LiDAR-stereo calibration results based on the provided left and right camera ground-truth calibration against other state-of-art targetless approaches, which are also evaluated on the KITTI dataset. Besides that, the co-calibration, which combines these two approaches, is also evaluated. We evaluate our uncertainty analysis of the estimated extrinsic parameters under some typical driving scenarios and explain the meaning of the estimated uncertainty. The evaluation shows that our approaches are accurate and robust by applying urban driving scenarios.

B. Results and Analysis

1) LiDAR-stereo calibration: To use the approach presented in Sec. III-B and Sec. III-C to extrinsic calibrate the LiDAR and stereo camera setup, a batch optimization with a batch size of 20 is applied. This procedure is processed as a sliding window approach over time of a driving sequence. To use measurements with meaningful information, we select measurements with a time interval of 0.5 s. The rotation initialization error is gaussian randomly simulated with a mean error of $3.0°$ and the translation initialization error with a mean error of 0.3 m. The hyperparameter to balance the point coordinates and normal vector ω in Eq.1 is chosen as 0.5. The calibration results based on an the KITTI sequence 00 are plotted in Fig. 5 and the comparison of our calibration results on the KITTI dataset against other state-of-the-art approaches are shown in TABLE. I.

As shown in Fig. 5, the coarse calibration is robust against bad initializations. After the coarse calibration, the rotation error of the calibration result is smaller than $0.6°$ and the translation error smaller than 0.2 m. After the fine calibration, the rotation error of the result is smaller than $0.1°$ and the translation error about 0.01 m. The estimated extrinsic parameters between the LiDAR and stereo camera have a lower covariance and our approach has a high consistency. Besides that, as shown in TABLE. I, our two-stage LiDAR-stereo calibration approach outperforms all of other state-of-the-art targetless extrinsic calibration approaches.

A calibration result example tested on the KITTI dataset is illuminated in Fig. 4. Fig. 4a shows the reprojection error

(a) The reprojection error in the image domain

(b) The distance error in the 3D space

Fig. 4: An LiDAR-stereo calibration error example tested on the KITTI dataset in the image domain and the 3D space. The green points and meshes denote the LiDAR points and triangle faces. The black points denote the stereo points.

on the image plane and both of the pillar edges are matched accurately. The calibration results are also shown in Fig.4b in the 3D space. The wall and vehicles in the LiDAR and stereo point clouds are aligned also accurately. This example proves that our approach achieves high-accurate results.

2) Left and right camera calibration: In this section, we evaluate the left and right camera calibration presented in Sec. IV-A and Sec. IV-B. We assume that the transformation between the LiDAR and stereo camera setup is known. As same as the LiDAR-stereo calibration, we apply a batch optimization with a batch size of 20 and the measurements are sampled with a time distance of 0.5 s. The initial rotation and translation error are gaussian randomly simulated with a mean errors of $1.0°$ and 0.1 m. The calibration results evaluated on the KITTI sequence 00 are shown in Fig. 5 in blue. Our approach achieves a high-level accurate performance. The most rotation error is smaller than $0.1°$ and the translation error smaller than 0.01 m.

3) Co-calibration: In this section, we evaluate the co-calibration process, which estimates the LiDAR-stereo extrinsic parameters and the left and right camera extrinsic parameters iteratively. Firstly, we start with the left and right camera calibration using the LiDAR depth information, because as explained in Sec. IV-C this approach is not sensitive to the key point depth noise. And then we run the LiDAR-stereo calibration with the estimated stereo camera calibration. Afterwards, the estimated LiDAR-stereo extrinsic parameters are fed into the left and right camera

TABLE I: Comparing our LiDAR-stereo calibration with other state-of-the-art approaches like HEC[16], STC[17], SOIC [18], MBC [19] and ECM [20] tested on the KITTI dataset. Our approach outperforms all of these approaches.

	Rx(°)	Ry(°)	Rz(°)	Tx(cm)	Ty(cm)	Tz(cm)
Ours	-0.002	-0.012	0.037	0.664	-0.513	0.331
HEC	0.800	-5.360	1.600	1.460	13.310	56.970
STC	-0.390	0.010	0.060	-2.150	-31.710	9.800
SOIC	0.070	-0.170	-0.230	6.100	-8.600	9.000
MBC	0.438	0.415	0.154	6.300	7.500	2.000
ECM	0.353	0.283	0.313	2.300	3.800	2.300
ECE	0.042	0.030	-0.040	-0.520	0.700	1.310

calibration process. This operation is repeated until both of these two calibration processes are converged. The batch optimization has a batch size of 10 and a sampling interval of 0.5 s. The initial rotation and translation error is gaussian randomly simulated with a mean errors of $1.0°$ and 0.1 m. The calibration results evaluated on the KITTI sequence 00 are shown in Fig. 7. For the left and right camera calibration, most of the rotation error is smaller than $0.1°$ and the translation error smaller than 0.02 m. For the LiDAR-stereo camera calibration, most of the rotation error is smaller than $0.3°$ and the translation error smaller than 0.05 m.

C. Uncertainty under Different Driving Scenarios

The targetless calibration approaches can provide extrinsic parameters automatically without any calibration targets. However, most of the targetless approaches have limitations on suitable driving scenarios. Using the uncertainty analysis, the estimation reliability of the extrinsic parameters under different scenarios can be obtained after optimization. We select 2 typical scenarios for the uncertainty analysis evaluation of our LiDAR-stereo calibration approach. The Fig. 8a and Fig. 8b denote a highway and an urban environment from the KITTI dataset. The calibration errors from the highway and urban environment are shown in Eq.14 and Eq. 15.

In the highway scenario, the translation error along the x- and z-axis as well as the rotation error along the y-axis are larger than the error in other dimensions. It is actually easy to understand. In this scenario, we try to align 2 planes. If these two planes have a big translation difference along the x- and z-axis and a big rotation difference along the y-axis, the residuals are still small. It means that the reliability of the calibrated translation result along the x- and z-axis and the rotation result along the y-axis is poor.

$$\mathbf{E_R} = \begin{pmatrix} 0.2116 & 1.8645 & 0.0398 \end{pmatrix} (°)$$
$$\mathbf{E_T} = \begin{pmatrix} -1.3658 & 0.0218 & -2.2419 \end{pmatrix} (m) \quad (14)$$

In the urban scenario, the error in each dimension is small, because the scenario provides enough constraints in each direction. Our algorithm achieves accurate and reliable calibration result by applying constraints from this scenario.

$$\mathbf{E_R} = \begin{pmatrix} 0.036 & -0.042 & 0.042 \end{pmatrix} (°)$$
$$\mathbf{E_T} = \begin{pmatrix} 0.004 & 0.015 & 0.006 \end{pmatrix} (m) \quad (15)$$
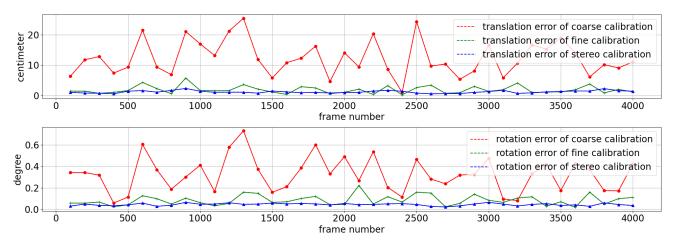
Fig. 5: The LiDAR-stereo as well as the left and right camera calibration results evaluated on the sequence 00 of the KITTI dataset. The red line shows the LiDAR-stereo coarse calibration result errors (gaussian randomly initialized with a rotation mean error of $3.0°$ and a translation mean error of 0.3 m). The blue line shows the left and right camera calibration result errors (gaussian randomly initialized with a rotation mean error of $1.0°$ and a translation mean error of 0.1 m).
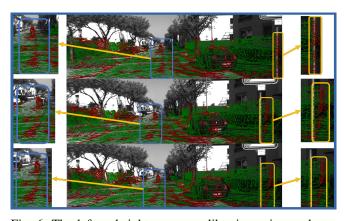

Fig. 6: The left and right camera calibration using a photometric error function. The upper image comes from the left camera. The middle and lower image come from the right camera before and after calibration. The key points in the blue rectangle are aligned accurately after the optimization.

The uncertainty ratio (the uncertainty under the highway scenario element wise divided by the uncertainty under the urban scenario) shown in Eq. 16 reflects the calibration reliability under these two driving scenarios.

$$\mathbf{\Sigma_R} = \begin{pmatrix} 0.705 & -120.293 & -0.308 \\ -120.293 & 5.593 & 91.499 \\ -0.308 & 91.499 & 0.062 \end{pmatrix}$$
$$\mathbf{\Sigma_T} = \begin{pmatrix} 12.364 & -9.272 & -13.966 \\ -9.272 & 0.728 & 2.215 \\ -13.966 & 2.215 & 3.989 \end{pmatrix} \quad (16)$$

The rotation uncertainty along the y-axis and the translation uncertainty along the x- and z-axis under the highway scenario are larger than the uncertainty under the urban environment. Because more LiDAR and stereo points on the ground under the highway scenario are used so that

the translation uncertainty along the y-axis and the rotation uncertainty along the x- and z-axis is smaller than the uncertainty under the urban environment. So the uncertainty analysis can provide us automatically the reliability of the estimated extrinsic parameters in all 6 DoF.

VIII. CONCLUSION AND FUTURE WORK

In this work, we present TEScalib, a targetless LiDAR and stereo camera extrinsic self-calibration approach especially for autonomous driving platforms. During the LiDAR-stereo calibration, a novel geometric information-based data association and objective function using point position and surface normal information are applied to estimate the extrinsic parameters accurately and robustly. The left and right camera calibration uses image key points and LiDAR depth to build a photometric objective function, which is then minimized to realize a highly accurate stereo camera extrinsic calibration. During driving, the so called co-calibration processes these two approaches iteratively. Besides, we propose an uncertainty analysis, which reflects the reliability of the estimated extrinsic parameters under different driving scenarios.

Our TEScalib makes it possible for autonomous driving platforms to extrinsic calibrate the LiDAR and stereo camera setup online robustly and accurately only using urban driving scenarios without any manufactured calibration targets from bad initializations. Using our uncertainty analysis, the quality of the estimated extrinsic parameters can be judged for different moving dimensions separately. Besides that, this automatic judgement also allows us to reject bad driving scenarios automatically to improve the calibration results.

For the future work we will address three points: realizing the LiDAR-stereo calibration and the left and right camera calibration in one optimization jointly; designing an online monitoring system to detect the calibration error and to decide whether a new calibration process should be triggered; analyzing of constraints for each single dimension.
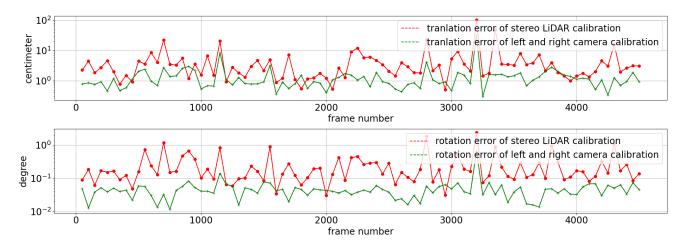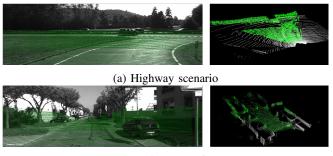
Fig. 7: The co-calibration results evaluated on the sequence 00 of the KITTI dataset. The red line denotes the LiDAR-stereo calibration result errors and the blue line the left and right camera calibration result errors. The initialization of the co-calibration is gaussian randomly simulated with a rotation mean error of $1.0°$ and a translation mean error of 10.0 cm).

(a) Highway scenario

(b) Urban scenario

Fig. 8: Uncertainty analysis under two scenarios. The stereo points are green colored and the LiDAR points white colored.

REFERENCES

[1] G. Pandey, J. McBride, S. Savarese, and R. Eustice, "Automatic Targetless Extrinsic Calibration of A 3d Lidar and Camera by Maximizing Mutual Information," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 26, 2012.

[2] M. Á. Muñoz-Bañón, F. A. Candelas, and F. Torres, "Targetless Camera-LiDAR Calibration in Unstructured Environments," *IEEE Access*, vol. 8, pp. 143 692–143 705, 2020.

[3] Z. Taylor and J. Nieto, "Motion-based Calibration of Multimodal Sensor Extrinsics and Timing Offset Estimation," *IEEE Transactions on Robotics*, vol. 32, no. 5, pp. 1215–1229, 2016.

[4] R. Ishikawa, T. Oishi, and K. Ikeuchi, "Lidar and Camera Calibration Using Motions Estimated by Sensor Fusion Odometry," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2018, pp. 7342–7349.

[5] C. Park, P. Moghadam, S. Kim, S. Sridharan, and C. Fookes, "Spatiotemporal camera-lidar calibration: A targetless and structureless approach," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1556–1563, 2020.

[6] K. Irie, M. Sugiyama, and M. Tomono, "Targetless Camera-Lidar Extrinsic Calibration Using a Bagged Dependence Estimator," in *2016 IEEE International Conference on Automation Science and Engineering (CASE)*, IEEE, 2016, pp. 1340–1347.

[7] J. Kang and N. L. Doh, "Automatic Targetless Camera–LIDAR Calibration by Aligning Edge With Gaussian Mixture Model," *Journal of Field Robotics*, vol. 37, no. 1, pp. 158–179, 2020.

[8] T. Cui, S. Ji, J. Shan, J. Gong, and K. Liu, "Line-based Registration of Panoramic Images and LiDAR Point Clouds for Mobile Mapping," *Sensors*, vol. 17, no. 1, p. 70, 2017.

[9] J. Jeong, Y. Cho, and A. Kim, "The Road Is Enough! Extrinsic Calibration of Non-Overlapping Stereo Camera and Lidar Using Road Information," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2831–2838, 2019.

[10] C. Shi, K. Huang, Q. Yu, J. Xiao, H. Lu, and C. Xie, "Extrinsic Calibration And Odometry for Camera-LiDAR Systems," *IEEE Access*, vol. 7, pp. 120 106–120 116, 2019.

[11] H. Hirschmuller, "Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, Jun. 2005, 807–814 vol. 2.

[12] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*. " O'Reilly Media, Inc.", 2008.

[13] R. Wang, M. Schworer, and D. Cremers, "Stereo dso: Large-scale direct sparse visual odometry with stereo cameras," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 3903–3911.

[14] C. Yuan, X. Liu, X. Hong, and F. Zhang, "Pixel-level extrinsic self calibration of high resolution lidar and camera in targetless environments," *arXiv preprint arXiv:2103.01627*, 2021.

[15] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision Meets Robotics: The Kitti Dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[16] H. Nguyen and Q.-C. Pham, "On the covariance of ax=xb," *IEEE Transactions on Robotics*, vol. 34, no. 6, pp. 1651–1658, 2018.

[17] C. Park, P. Moghadam, S. Kim, S. Sridharan, and C. Fookes, "Spatiotemporal camera-lidar calibration: A targetless and structureless approach," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1556–1563, 2020.

[18] Z. Xiao, H. Li, D. Zhou, Y. Dai, and B. Dai, "Accurate Extrinsic Calibration Between Monocular Camera and Sparse 3D Lidar Points without Markers," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, 2017, pp. 424–429.

[19] W. Wang, S. Nobuhara, R. Nakamura, and K. Sakurada, "SOIC: Semantic Online Initialization and Calibration for LiDAR and Camera," *arXiv preprint arXiv:2003.04260*, 2020.

[20] Z. Taylor and J. Nieto, "Motion-based Calibration of Multimodal Sensor Extrinsics and Timing Offset Estimation," *IEEE Transactions on Robotics*, vol. 32, no. 5, pp. 1215–1229, 2016.