Road is Enough! Extrinsic Calibration of Non-overlapping Stereo Camera and LiDAR using Road Information

Jinyong Jeong[1], Lucas Y. Cho[1] and Ayoung Kim[1*]

Abstract— This paper presents a framework for the target-less extrinsic calibration of stereo cameras and Light Detection and Ranging (LiDAR) sensors with a non-overlapping Field of View (FOV). In order to solve the extrinsic calibrations problem under such challenging configuration, the proposed solution exploits road markings as static and robust features among the various dynamic objects that are present in urban environment. First, this study utilizes road markings that are commonly captured by the two sensor modalities to select informative images for estimating the extrinsic parameters. In order to accomplish stable optimization, multiple cost functions are defined, including Normalized Information Distance (NID), edge alignment and, plane fitting cost. Therefore a smooth cost curve is formed for global optimization to prevent convergence to the local optimal point. We further evaluate each cost function by examining parameter sensitivity near the optimal point. Another key characteristic of extrinsic calibration, repeatability, is analyzed by conducting the proposed method multiple times with varying randomly perturbed initial points.

I. INTRODUCTION

Autonomous vehicles are often equipped with several multi-modal sensors for environmental perception. Typical perceptual sensors include range sensors which measure the distance to objects using signals (e.g., light, radio waves, and sound waves) and vision sensors which project 3D information onto a 2D image frame. If multiple sensor types are equipped, obtaining relative coordinate transformations via extrinsic calibration becomes an essential pre-processing module. According to literature, accurate extrinsic calibration enhances the performance of data fusion between multiple sensors and the complementary exploitation of other modalities. For example, Shin et al. [1] utilized depth information from LiDAR to calculate the photometric error of two images. Zhang and Singh [2] implemented a fast odometry method featuring low drift by fusing visual odometry and LiDAR odometry methods. Han et al. [3] conducted a study to estimate road area using LiDAR data and camera images using a conditional random field framework.
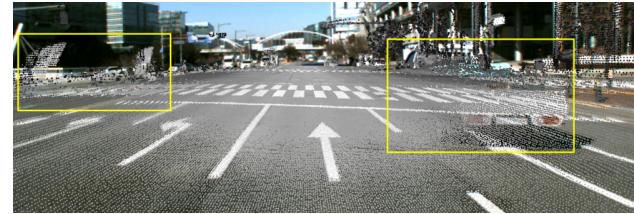
The most common extrinsic calibration method between LiDAR and cameras involves the use of a checkerboard-like target and the process of minimizing the re-projection error of the correspondences between two sensor data [4, 5, 6]. Some studies adopted a modified target such as a circular hole [7, 8]. However, such calibration methods, require a

(a) Pointcloud data from LiDAR

(b) Projected pointcloud onto a stereo image using calibration result.

Fig. 1: Result of extrinsic calibration between a stereo camera and LiDARs. The projected pointcloud in the image plane coordinate system presents well-matched sensor modalities. The yellow boxes show the inconsistency of the data due to the time difference in the data acquisition.

target and the performance of the calibration depends on the accuracy of the correspondence estimated from each sensor data. In particular when LiDAR measurements are non-continuous, the detection of an accurate correspondence between the two modalities is not straightforward. Furthermore, this type of calibration is not possible if the two sensors do not provide covisibility at a given moment in time. For the purpose of general calibration, several algorithms were introduced to perform calibration using information on the surrounding environment from sensors without using a specific target [9, 10, 11, 12, 13]. However, these approaches rely on an overlapping FOV between two sensors. In order to overcome the non-overlapping configuration issue, a local 3D pointcloud is accumulated using the odometry to ensure that the overlapping region appears on the camera [14, 15, 16]. Another line of study [17, 18, 19], referred to as hand-eye calibration, estimates the motion of each sensor and performs extrinsic calibration based on the estimated motion relationship.

The configuration of our sensor system for extrinsic calibration consist of four LiDARs and a single stereo camera, as shown in Fig. 2. Two 3D LiDARs are mounted on the left and right to maximize the range of data acquisition. The two 2D LiDARs are installed facing forward and backward, respectively. As the four LiDARs have overlapping regions,
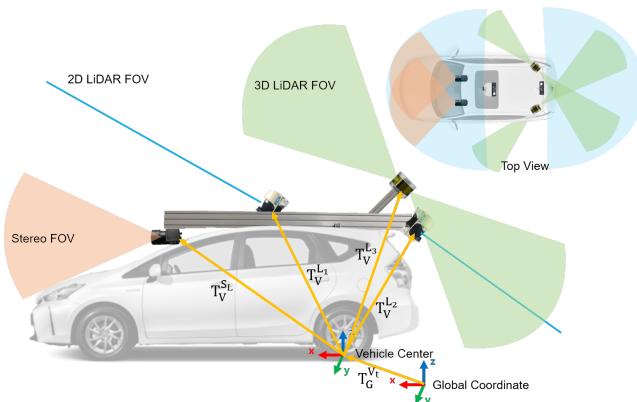
Fig. 2: System configuration of the target mobile mapping system. Extrinsic calibration between stereo cameras and LiDAR sensors is a challenging task when there is no guaranteed covisibility.

calibration was performed using the geometric information from LiDARs. However, as the stereo camera is facing forward, the images from the stereo camera possess no overlap with any of the LiDAR sensor data.

II. RELATED WORKS

In this paper, we perform an extrinsic calibration between a stereo camera and LiDARs when no overlapping FOV is guaranteed. Literature sources related to our extrinsic calibration method can be divided into two major categories: target-less and calibration with non-overlapping configurations.

A. Target-less Extrinsic Calibration

Target-less extrinsic calibration refers to the method of performing calibration using surrounding data from a general space without the use of a checkerboard or a specific shape target. Establishing correspondences between heterogeneous sensor data without a specific target is a highly challenging task.

Zhao et al. [12] manually selected the correspondence between LiDAR intensity image and camera image to calculate the extrinsic parameter. LiDAR intensity image was generated by fusing LiDAR scan data and odometry estimated using horizontal 2D LiDAR data. However, the selection process for correspondence between heterogeneous sensor data is a laborious task and potential source of human errors. In order to mitigate this issue, many researchers have attempted to calibrate without correspondence through the use of information theory. Taylor and Nieto [9] proposed to find the optimal alignment between sensors by introducing the Normalized Mutual Information (NMI) metric. NMI is a measure of the mutual dependence of data from two sensors. This study yields the optimal solution by finding the largest NMI value using the particle swarm optimization method. Similarly, Pandey et al. [10] used the Mutual Information (MI) metric to estimate the extrinsic parameter. Several local optimal points are encountered if a single scan is used for optimization. In order to cope with the problem, the method

of this study used multiple scan data to find global optimal points with maximum MI value. Recently, with more studies using deep learning, there have been attempts to perform extrinsic calibration using deep learning technique. Iyer et al. [11] used the CNN architecture to estimate calibration parameters using a supervised method.

B. Extrinsic Calibration of Non-overlapping Configuration

The aforementioned target-less extrinsic calibration still requires an overlap between sensors. Several studies have focused on solving extrinsic calibration for configurations without any overlap between sensors. For example, Napier et al. [16] performed extrinsic calibration between push-broom 2D LiDAR and cameras. The push-broom LiDAR obtained distance measurement from the left-hand and the right-hand sides of the vehicle, and the cameras obtained images from the front. This means that there was no overlapping region between the two sensor data. Therefore, extrinsic calibration was performed using edge images from the camera image and LiDAR intensity image generated by projecting the 3D pointcloud that was calculated by applying the vehicle odometry measurement.

Scott et al. [14] leveraged the Normalized Information Distance (NID) metric to estimate optimal extrinsic parameter between 2D LiDAR and multiple cameras. In the optimization process, they utilized factory-calibrated transformation as constraint. Scott et al. [15] added a data selection module by comparing the NID value around the optimal parameter to reduce the local optimal points.

Our proposed method utilizes the road marking information for the extrinsic calibration of non-overlapping configurations. The road are observed together through both sensors (camera and LiDAR) regardless of the acquisition time of the data. In order to minimize the local optimal point, multiple images were used and the images were selected using the costs defined with the vanishing point. In the optimization process, we defined multiple costs using image and LiDAR data which include NID cost, edge alignment cost and plane fitting cost to make it more convergent. This paper presents the following:

- An automatic extrinsic calibration framework for non-overlapping configuration
- Utilization of road regions estimated using stereo images for a more stable optimization process
- An automatic image selection process using estimated vanishing point and road markings
- A multiple cost function for robust optimization even with rough initial values

III. NOTATION AND LOCAL MAP GENERATION

We first generated a 3D pointcloud using odometry and LiDAR data and performed calibration by comparing the stereo images and LiDAR intensity image over the generated local pointcloud map. For local map generation, the odometry of the vehicle was calculated with high accuracy using the wheel encoder, the 3-axis Fiber Optic Gyro (FOG), and the Inertial Measurement Unit (IMU). Before describing the
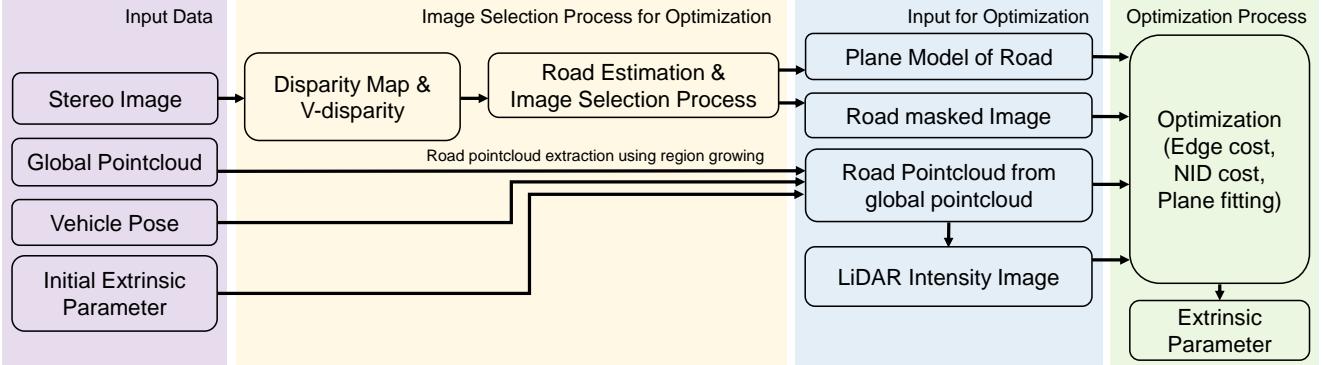
Fig. 3: Overall process of automatic calibration. The colored boxes refers to the different processes. Details of the image selection process are described in Fig 4.

details of the extrinsic calibration, this paper first introduces notations and assumptions for generating the local map.

A. Notation

Matrix $T_B^A \in SE(3)$ represents the rigid body transformation that registers the data defined in A coordinate system to B coordinate system. Matrix T_B^A is also represented by a tuple $t_B^A \in \mathbb{R}^6$, where $t_B^A = \{t_x, t_y, t_z, r_x, r_y, r_z\}$. Here, t_x, t_y and t_z represent the relative translation along each axis in meters, and r_x, r_y and r_z represent the relative rotation that is roll, pitch, and yaw in radians. $^C P_D$ represents the pointcloud set from sensor D in coordinate system C. For instance, $^{V_t} P_{L_{i,t}}$ represents the pointcloud from the i_{th} LiDAR at time t in the coordinate system V. Subscripts $L, S_L, V,$ and G denote LiDAR, the left camera of stereo, the vehicle center and the global coordinate system.

B. Local Map Generation

Due to the fact that no overlap is attained between the camera and the LiDAR at any given instance of time, we accumulated LiDAR data to generate a local pointcloud map. We assume that accurate relative transformation is calculable using a 3-axis FOG, an IMU and wheel encoder. We also began with pre-computed extrinsic calibration between Li-DARs and LiDAR-to-vehicle using our previously proposed algorithm [20].

Given an accurate relative sequential pose and extrinsic calibration between the four LiDARs, we accumulated the LiDAR data and constructed a local pointcloud map written in the global coordinate system. The local pointcloud $^{L_i} P_{L_{i,t}}$ that was written in each sensor coordinate system at time t was represented in the vehicle coordinate system using the computed vehicle-to-LiDAR coordinate system transform $t_V^{L_i}$ as described in (1). This was followed by the sequential transformation of pointcloud $^{V_t} P_{L_{i,t}}$ from each LiDAR sensor in the vehicle coordinate system into the global coordinate frame via (2). Lastly, the global pointcloud $^G P_L$ was obtained by adding of the all pointcloud of each LiDAR via (3).

Our method used the global pointcloud $^G P_L$, the stereo image I_{S_L}, and the odometry $t_G^{V_t}$ for stereo extrinsic calibration. Particular attention was required as the accuracy of the

local pointcloud map depends on the accuracy of the relative vehicle pose estimation. In this study, we accumulated the local map for a range of approximately $80\,\text{m}$ to minimize the potential accumulation error.

$$^{V_t} P_{L_{i,t}} = t_V^{L_i} \oplus {}^{L_i} P_{L_{i,t}} \tag{1}$$

$$^{G} P_{L_{i,t}} = t_G^{V_t} \oplus {}^{V_t} P_{L_{i,t}} \tag{2}$$

$$^{G} P_L = \sum_{i=1}^{n} \sum_{t=0}^{T} {}^{G} P_{L_{i,t}} \tag{3}$$

The local pointcloud map was obtained at time t when the LiDAR measurement was received. As the camera images were captured at different frequencies, we needed to consider the local pointcloud at time $t + k$ when the image is received. The pointcloud in the global coordinate system can be transformed into the vehicle coordinate system using the odometry information $t_{V_{t+k}}^G$ at time $t + k$ time at which the stereo image was acquired (4). The pointcloud $^{V_{t+k}} P_L$ in the vehicle coordinate system was subsequently registered into the stereo camera coordinate system using the extrinsic parameter $t_{S_L}^V$ of the stereo camera, which is what our method is aiming to calculate. (5). The LiDAR intensity image was generated by projecting the pointcloud onto stereo image plane (Fig. 1) using the pre-calibrated intrinsic parameter K_{S_L} (6).

$$^{V_{t+k}} P_L = t_{V_{t+k}}^G \oplus {}^{G} P_L \tag{4}$$

$$^{S_{L,t+k}} P_L = t_{S_L}^V \oplus {}^{V_{t+k}} P_L \tag{5}$$

$$^{S_{L,t+k}} I_L = K_{S_L} {}^{S_{L,t+k}} P_L \tag{6}$$

IV. TARGET-LESS MULTI-MODAL EXTRINSIC CALIBRATION WITH NO OVERLAP

With the constructed local map, the proposed method peforms two steps: (i) informative image selection and (ii) multiple cost optimization. In the image selection process, the road regions are extracted using a stereo disparity map, and images that include abundant road markings were selected for the optimization process by extracting the lines facing the vanishing point. The local optima issue can be alleviated by using multiple images to estimate extrinsic
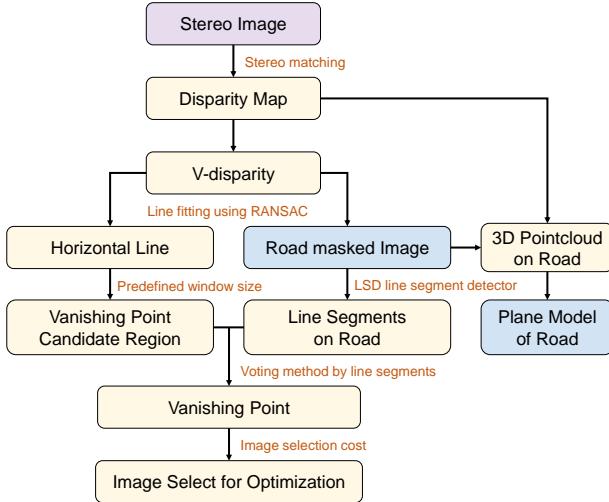
Fig. 4: Image selection process using stereo images.

(a) Left image of the stereo camera and v-disparity

(b) Disparity map $D(p_{uv})$ of stereo image

(c) Road mask extracted by plane fitting in v-disparity

Fig. 5: Road extraction process using a stereo image. The disparity map is estimated via stereo matching and the v-disparity is calculated through projection onto the plane of (v, disparity). The π in the v-disparity refers to the road surface.

calibration. Fig. 3 shows the overall calibration process. The following sections describe each module in detail.

A. Road Detection and Informative Image Selection

Images are marked as informative if abundant road markings are contained. In this section, we detect the road region within an image and propose the image utility measure for extrinsic calibration. The overall scheme is provided in Fig. 4.

1) Road Region Detection using Stereo Images: In the proposed method, we compute the v-disparity image [21] from a disparity map $D(p_{uv})$ for a pixel p_{uv} to detect the road region. The histogram of the row line on the disparity map is depicted on the row of the v-disparity image, and sample v-disparity images are shown on the right-hand side of Fig. 5.
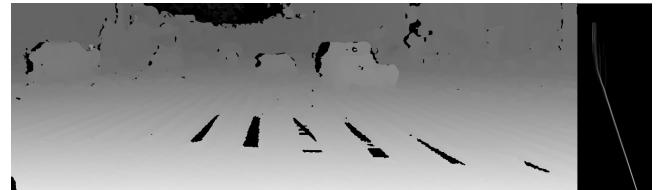
Following this process, the road region is detected by fitting a line π on the v-disparity image using Random Sample Consensus (RANSAC). Due to the fact that each pixel value of the v-disparity image denotes the histogram of the disparity map, fitting a line captures the continuously changing disparity along the v direction (i.e., ground) of the image. The road region is informed from pixels corresponding to the fitted line π. The sample of the detected road region can be seen in Fig. 5(c).

2) Vanishing Point Estimation: Next, the vanishing point is estimated from the detected road region. The vanishing point allows for the discerning of road markings from other features such as the shadows of street lamps and trees. Given the line π in the v-disparity image (i.e., the road plane), the v value at which π meets the v- axis is identified. The horizontal line at this v value is referred to as the $L_{horizon}$, as depicted in Fig. 5(a).
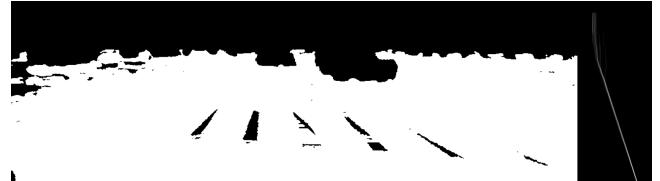
The $L_{horizon}$ is near to the vanishing point but may not be exact. A voting process is additionally applied to estimate the vanishing point near the $L_{horizon}$. After defining a voting region R_{van} of size W_w by W_h around the central point of $L_{horizon}$, the lines $\widehat{L_{S_L}} = \{l_1, l_2, \cdots, l_n\}$, which are voters

participating in the vote, were detected on the road region using a Line Segment Detector (LSD) [22]. The symbol $\widehat{\cdot}$ is used to indicate that the corresponding data was extracted from the road region. Each line $l_i = \{s_i, e_i, c_i\}$, where s_i, e_i and c_i are the start, end and center point of each line respectively, votes depending on the weight over each pixel p_{uv} in the R_{van} (7).

$$\text{Vote}(p_{uv}) = \begin{cases} \min\left(\frac{1}{\alpha}, 10\right) & \text{if } \alpha \leq \rho_1 \\ 0 & otherwise \end{cases} \quad (7)$$
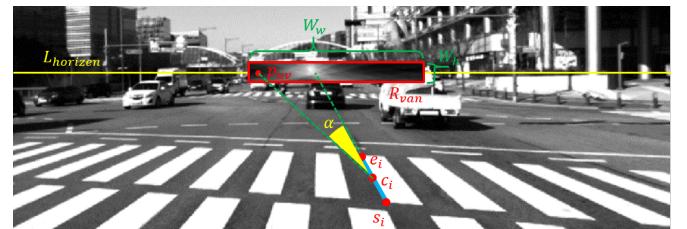

Fig. 6: Voting process for vanishing point estimation. The yellow line refers to the horizon calculated using v-disparity and the red box refers to the candidate vanishing point region. Each line extracted from the road image votes on every pixel in the candidate area.
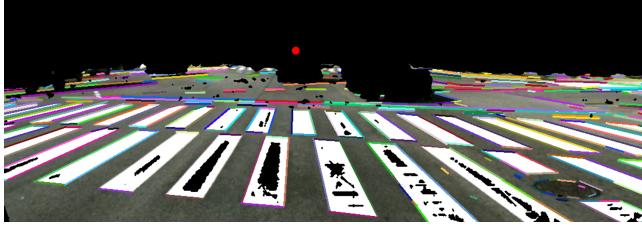
Fig. 7: Result of the vanishing point extraction process. The red point represents the estimated vanishing point. The lines extracted from the road are voters that participate in the vote.

$$p_{van} = \underset{p_{uv}}{\arg\max} \, R_{van}(p_{uv}) \qquad (8)$$

$$U_{van} = \max R_{van}(p_{uv}) \qquad (9)$$

Here, $\alpha = \angle(\overline{p_{uv}, c_i}, \overline{s_i, e_i})$ refers to an angle between voter l_i and the line connecting c_i and p_{uv} in degrees, as illustrated in Fig. 6. The voting weight of each line is the inverse of the angle to emphasize lines with smaller angles (with a maximum weight of 10). In this paper, threshold ρ_1 was set to $3°$. The vanishing point p_{van} was determined as the largest value in the voting results (8), and U_{van} denotes the corresponding confidence level. Fig. 7 shows the estimated vanishing point and line segments on the road.
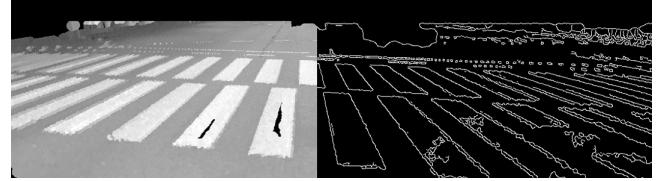
3) Image Selection: The utility of each image U_I for image selection is calculated using p_{van}, $\widehat{L_{S_L}}$ and U_{van} (10). A higher U_I represents a greater number of road markings, which in turn indicates better data for calibration. In order to suppress the effect of the local optimal point in the optimization process, multiple images and the corresponding pointcloud were selected. Based on the utility, five images with the highest utility are selected.

$$U_I = \left(\sum_{i=1}^{N} \min\left(\frac{1}{\angle(\overline{p_{van}, c_i}, \overline{s_i, e_i})}, 1\right)\right) \cdot U_{van} \qquad (10)$$
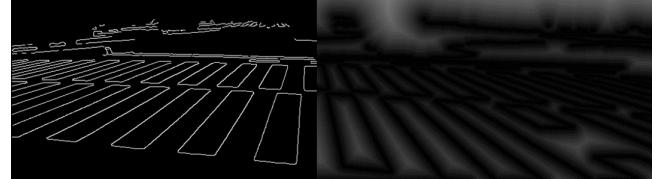
4) Plane Estimation of the Road: Once the informative images are selected, the plane $M = \{n_x, n_y, n_z, d\}$, is estimated from the disparity map $D(p_{uv})$. Conversions are made from $D(p_{uv})$ in camera coordinate system to the 3D pointcloud using the camera model: the focal length f, and base line B. M is subsequently calculated using a RANSAC plane fitting with the pointcloud. The plane is used in optimization process to calculate the plane fitting cost.

B. Multi-cost Optimization

Through the aforementioned image selection process, a LiDAR intensity image, a road masked image, and a plane model were obtained. For the extrinsic calibration, we also extracted points belonging to roads. Specifically, the road pointcloud $^{S_L}\widehat{P_L}$ was calculated from the global pointcloud using the region growing segmentation method [23]. For extrinsic calibration, this paper proposes using the following three costs.

(a) The LiDAR intensity image $\widehat{I_L}$ (left) and edge image $\widehat{E_L}$ extracted from LiDAR intensity image (right)

(b) Edge image $\widehat{E_{S_L}}$ extracted from stereo image (left) and distance transform image $\widehat{G_{S_L}}$ using inverse of edge image from stereo (right)

Fig. 8: The process of computing the edge alignment cost. The edge is extracted using the canny edge algorithm with the LiDAR intensity image and the stereo image. The distance transformed image is calculated using the inversely transformed edge image of the stereo image.

1) Edge Alignment Cost: The first cost, edge alignment cost, is used to evaluate the discrepancy between RGB and LiDAR intensity images. A camera edge image ($\widehat{E_{S_L}}$) is extracted from a road masked image ($\widehat{I_{S_L}}$), and a LiDAR edge image ($\widehat{E_L}$) is computed from a LiDAR intensity image ($\widehat{I_L}$) by applying the canny edge algorithm [24].

In order to compare the two edge images by defining a differentiable cost function, RGB-induced edge image is converted into a distance transform image $\widehat{G_{S_L}}$. The edge alignment cost was computed by calculating the sum of the multiplied pixel value of $\widehat{E_L}$ and $\widehat{G_{S_L}}$ (11). Defining the cost using the distance image aids the algorithm to converge with rough initial parameters. The edge cost reaches a minimum value when the two data precisely match one another. Fig. 8 shows the process of calculating the edge alignment cost.

$$f_{edge}(\widehat{E_L}, \widehat{G_{S_L}}) = \sum_{p_{uv}} \widehat{E_L}(p_{uv})\widehat{G_{S_L}}(p_{uv}) \qquad (11)$$

2) NID Cost: The second cost is induced from the NID metric, which measures the correlation between two random variables, X and Y. In this paper, the NID cost between $\widehat{L_{S_L}}$ and $\widehat{I_L}$ was used to identify the best alignment. The NID cost was defined as:

$$f_{NID}(\widehat{I_{S_L}}, \widehat{I_L}) = 2 - \frac{H(\widehat{I_{S_L}}) + H(\widehat{I_L})}{H(\widehat{I_{S_L}}, \widehat{I_L})} \qquad (12)$$

$$H(X) = -\sum_{x \in X} P(x) \log P(x)$$

$$H(X, Y) = -\sum_{\substack{x \in X \\ y \in Y}} P(x,y) \log(P(x,y)).$$

Here, $H(X)$ and $H(X,Y)$ are single entropy and cross entropy. The computed NID metric has the range of $0 \leq$
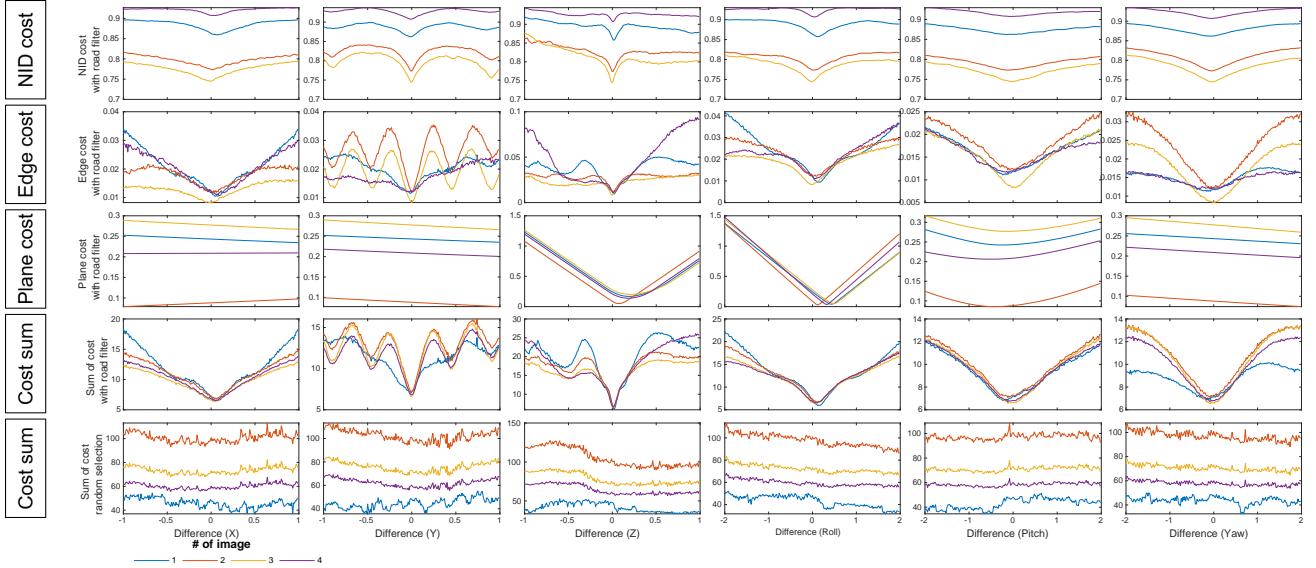
Fig. 9: Variation of each cost according to parameter changes based on the optimal extrinsic parameter. The first row to the fourth row shows the change in cost according to the number of images for the selected image and road filtering. The fifth row shows the cost change when the images are selected randomly. The units for translation and rotation are meter and degree respectively.

$f_{NID}(\widehat{I_{S_L}}, \widehat{I_L}) \leq 1$. Lower NID costs indicate a greater degree of similarity between the distributions of two data. In our case, the intensity value of LiDAR and the image pixel value in grayscale achieves the best alignment between two images with the lowest NID cost.

3) Plane Fitting Cost: The final cost is computed using the plane model M obtained from the stereo image and the road pointcloud $^{S_L}\widehat{P_L}$ acquired from the global 3D map. By including this geometric cost, convergence of the z and the rotation parameter is improved.

$$f_{plane}(^{S_L}P_L, M) = \sum_{i=1}^{N}\left(n_x(^{S_L}P_L)^i_x + n_y(^{S_L}P_L)^i_y + n_z(^{S_L}P_L)^i_z + d\right) \tag{13}$$

In this cost, $(^{S_L}P_L)^i_x$, $(^{S_L}P_L)^i_y$ and $(^{S_L}P_L)^i_z$ are the x, y, and z values of ith point in the stereo camera coordinate system. The cost is summed over the number of the points (N) in the target pointcloud $^{S_L}P_L$. Lower plane fitting costs are yielded when the pointcloud and road model closely match one another.

4) Optimization: Optimization is performed using the weighted sum of the three costs: f_{edge}, f_{NID} and f_{plane}. In this paper, k_1, k_2, and k_3 are set as 2.0, 500.0, 0.1 respectively, considering the scale of each cost.

$$f_{sum} = k_1 f_{edge} + k_2 f_{NID} + k_3 f_{plane} \tag{14}$$

$$t^V_{S_L} = \underset{t^V_{S_L}}{\operatorname{argmin}} f_{sum} \tag{15}$$

The extrinsic parameter of the stereo camera $t^V_{S_L}$ was estimated by identifying the minimum point of sum of the costs by utilizing the downhill simplex method (15).

V. EXPERIMENTAL RESULTS

This section presents experimental results to evaluate the defined cost in both a quantitative and qualitative manner. We also verified the repeatability and sensitivity of the extrinsic parameter between the stereo camera and LiDAR during the proposed extrinsic calibration.

A. Evaluation of Cost Function

We started by depicting the cost function with respect to the perturbation from the optimal values. The variation of each cost and total cost are plotted in Fig. 9 by increasing the number of selected images. From the first row to the fourth row, up to four informative images were incrementally included in the information order, from the most informative to the fourth most informative image. For these cases, the cost was calculated using only the information on the road region.

As shown in the results, the proposed cost reshapes the function in all parameter spaces. The NID cost has a smooth curve shape compared to the edge cost, whereas the segment for converging to the global optimal point is relatively narrow. On the other hand, the edge cost graph is somewhat noisier but has a broader segment to converge to the global optimal point. This feature of the edge cost helps to converge to the global optimal point even with a large initial error of calibration. The shape of the edge cost graph in the y-direction (lateral direction) can be explained considering that the selected informative images tended to contain more road

	X [m]	Y [m]	Z [m]	Roll [degree]	Pitch [degree]	Yaw [degree]
Left	1.669	0.278	1.612	-91.124	-0.632	-90.236
Right	1.710	-0.201	1.595	-90.985	-0.603	-90.146
Diff	0.041	0.479	0.017	0.139	0.029	0.09
GT	0	0.475	0	0	0	0
Error	0.041	0.004	0.017	0.139	0.029	0.09

TABLE I: Quantitative evaluation using a stereo camera. There should be a translation of the baseline in exactly the y-axis direction between the rectified images.

markings with repeated patterns (e.g., cross walks). However, note that the NID cost alleviated this repeated local minimum in the y-axis direction in the summed cost function.

The plane fitting cost affects z, roll, and pitch by exploiting the geometric road information, and mainly enhances the convergence in z and roll. The plane fitting cost depends on the accuracy of the disparity map. The computed disparity map may include errors due to pixel discontinuity, thus resulting in a less sharp shape at the optimal point for the optimal point. To incorporate the global tendency while hindering the local details of the plane cost, we set the weight of the plane fitting cost to be somewhat smaller. Overall, the sum of the cost shows a distribution of shapes that can converge toward the global optimal point in the overall section.

The graph in the last row reveals the effect of the image selection. The plot shows the cost variation without an image selection process, by selecting a image randomly. When an image is arbitrarily selected, the cost produces significant amount of noise for all the parameters and fails to provide a clear global optimal point regardless of the number of images. This proves that an informative image selection process significantly influences extrinsic parameter estimation.

B. Repeatability of Parameter Optimization

Repeatability is the property of outputting the same result for different inputs through repetition of the algorithm. In order to test repeatability, the algorithm was executed 40 times for each number of images using random initial values within the range of 0.3 m and $3°$ for translation and rotation, respectively (Fig. 10). The top and bottom of the box plot refer to the 25th and 75th percentiles respectively, and the middle red line is the median of the error. The higher the number of images, the better the overall repeatability as a whole, and the repeatability is significantly improved when two or more images are used.

C. Comparison with Available Ground Truth

It is challenging to calculate the ground truth of the actual calibration parameters for extrinsic calibration, especially when no overlap is guaranteed as in the case of our system. Therefore, in this paper, we used the baseline of the stereo camera as the ground truth. If the stereo camera is correctly calibrated, there should be a translation of the baseline in exactly the y-axis direction between the rectified images, and the rest of the translation and rotation should be zero.
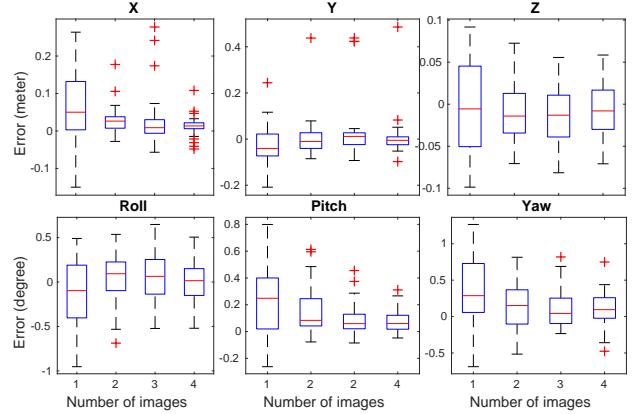

Fig. 10: Repeatability of the parameter estimation with respect to the number of images used. The calibration was performed 40 times using random initial points while measuring error against the calibration result of using all five informative images.

The extrinsic parameters of each of the two rectified images were calculated using the proposed algorithm, and the difference between the two parameters was compared. Table I shows the comparison with the ground truth. On average, there was an error of approximately 0.02 m for translation and approximately $0.086°$ for rotation. Fig. 11 shows the image when the global pointcloud is projected onto the left stereo camera image using the finally computed extrinsic parameter. We can see that the pointcloud of the road and the structure is projected correctly onto the corresponding pixels.

VI. CONCLUSION AND FUTURE WORKS

This paper proposed an automatic extrinsic calibration method for estimating the rigid body transformation between a stereo camera and LiDAR with no overlapping FOV. By exploiting the static features in the urban environment such as road markings, it was confirmed that extrinsic parameters
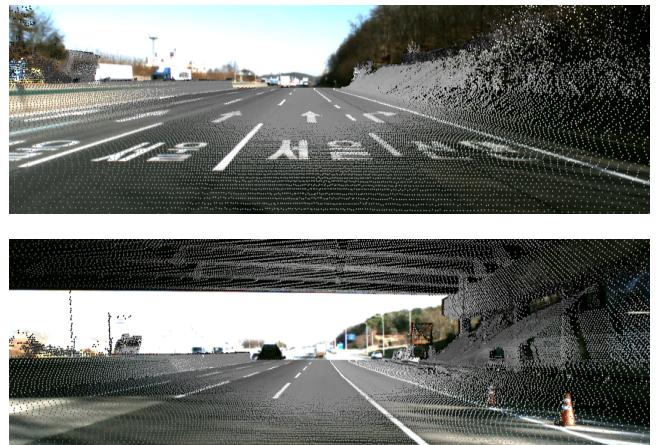

Fig. 11: Qualitative evaluation of the calibration. We projected the intensity value of the pointcloud onto the stereo image in a highway environment.

could be calculated with a degree of accuracy of approximately 0.02 m and 0.086° without manual operation by a human operator.

In this paper, we used the assumption that the odometry for generating poitncloud is locally accurate. However, in cases where the motion of a vehicle undergoes several accelerations and decelerations, there may be a distortions in the 3D map due to errors in motion estimation. Therefore, it is preferable to avoid making assumptions about the accuracy of the odometry to use the proposed method in all situations. Therefore, if the uncertainty of the odometry and the SLAM framework are included in the calibration framework, these two problems can be simultaneously solved together through the optimization process.

REFERENCES

[1] Y.-S. Shin, Y. S. Park, and A. Kim, "Direct visual SLAM using sparse depth for camera-lidar system," in *Proc. IEEE Intl. Conf. on Robot. and Automat.*, 2018, pp. 1–8.

[2] J. Zhang and S. Singh, "Visual-lidar odometry and mapping: Low-drift, robust, and fast," in *Proc. IEEE Intl. Conf. on Robot. and Automat.*, 2015, pp. 2174–2181.

[3] X. Han, H. Wang, J. Lu, and C. Zhao, "Road detection based on the fusion of lidar and image data," *Intl. J. of Robot. Research*, vol. 14, no. 6, pp. 1–10, 2017.

[4] K. M. A. Yousef, B. J. Mohd, K. Al-Widyan, and T. Hayajneh, "Extrinsic calibration of camera and 2D laser sensors without overlap," *IEEE Sensors J.*, vol. 17, no. 10, pp. 1–24, 2017.

[5] L. Zhou, "A new minimal solution for the extrinsic calibration of a 2D lidar and a camera using three plane-line correspondences," *IEEE Sensors J.*, vol. 14, pp. 442–454, 2014.

[6] A. K. Krishnan and S. Saripalli, "Cross-calibration of RGB and thermal cameras with a lidar for rgb-depth-thermal mapping," *Unmanned Sys.*, vol. 5, no. 02, pp. 59–78, 2017.

[7] J. Peršić, I. Marković, and I. Petrović, "Extrinsic 6DoF calibration of 3D LiDAR and radar," in *European Conf. on Mobile Robots*, 2017, pp. 1–6.

[8] H. Alismail, L. D. Baker, and B. Browning, "Automatic calibration of a range sensor and camera system," in *Proc. IEEE Intl. Conf. on 3D Imaging, Modeling, Processing, Visualization and Transmission*, 2012, pp. 286–292.

[9] Z. Taylor and J. Nieto, "Automatic calibration of Li-DAR and camera images using normalized mutual information," in *Proc. IEEE Intl. Conf. on Robot. and Automat.*, 2013.

[10] G. Pandey, J. R. McBride, S. Savarese, and R. M. Eustice, "Automatic extrinsic calibration of vision and lidar by maximizing mutual information," *J. of Field Robot.*, vol. 32, no. 5, pp. 696–722, 2015.

[11] G. Iyer, R. K. Ram, J. K. Murthy, and K. M. Krishna, "Calibnet: Geometrically supervised extrinsic calibra-tion using 3D spatial transformer networks," in *Proc. IEEE/RSJ Intl. Conf. on Intell. Robots and Sys.*, 2018, pp. 1110–1117.

[12] H. Zhao, Y. Chen, and R. Shibasaki, "An efficient extrinsic calibration of a multiple laser scanners and cameras' sensor system on a mobile platform," in *Proc. IEEE Intell. Vehicle Symposium*, 2007, pp. 422–427.

[13] L. Tamas and Z. Kato, "Targetless calibration of a lidar-perspective camera pair," in *Proc. of the IEEE Int. Conf. on Comp. Vision Workshops*, 2013, pp. 668–675.

[14] T. Scott, A. A. Morye, P. Piniés, L. M. Paz, I. Posner, and P. Newman, "Exploiting known unknowns: Scene induced cross-calibration of lidar-stereo systems," in *Proc. IEEE/RSJ Intl. Conf. on Intell. Robots and Sys.*, 2015, pp. 3647–3653.

[15] ——, "Choosing a time and place for calibration of lidar-camera systems," in *Proc. IEEE Intl. Conf. on Robot. and Automat.*, 2016, pp. 4349–4356.

[16] A. Napier, P. Corke, and P. Newman, "Cross-calibration of push-broom 2D lidars and cameras in natural scenes," in *Proc. IEEE Intl. Conf. on Robot. and Automat.*, 2013, pp. 3679–3684.

[17] Z. Taylor and J. Nieto, "Motion-based calibration of multimodal sensor extrinsics and timing offset estima-tion," *IEEE Trans. Robot.*, vol. 32, no. 5, pp. 1215–1229, 2016.

[18] N. Andreff, R. Horaud, and B. Espiau, "Robot hand-eye calibration using structure-from-motion," *Intl. J. of Robot. Research*, vol. 20, no. 3, pp. 228–248, 2001.

[19] J. Heller, M. Havlena, A. Sugimoto, and T. Pajdla, "Structure-from-motion based hand-eye calibration us-ing l∞ minimization," in *Proc. IEEE Conf. on Comput. Vision and Pattern Recog.*, 2011, pp. 3497–3503.

[20] J. Jeong, Y. Cho, Y.-S. Shin, H. Roh, and A. Kim, "Complex urban LiDAR data set," in *Proc. IEEE Intl. Conf. on Robot. and Automat.*, Brisbane, May. 2018, pp. 6344–6351.

[21] R. Labayrade, D. Aubert, and J.-P. Tarel, "Real time obstacle detection in stereovision on non flat road ge-ometry through" v-disparity" representation," in *Proc. IEEE Intell. Vehicle Symposium*, vol. 2. IEEE, 2002, pp. 646–651.

[22] R. G. V. Gioi, J. Jakubowicz, J.-M. Morel, and G. Ran-dall, "LSD: A fast line segment detector with a false detection control," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 32, no. 4, pp. 722–732, 2010.

[23] T. Rabbani, F. V. D. Heuvel, and G. Vosselmann, "Seg-mentation of point clouds using smoothness constraint," *Intl. archives of photogrammetry, remote sensing and spatial info. sci.*, vol. 36, no. 5, pp. 248–253, 2006.

[24] J. Canny, "A computational approach to edge detec-tion," *IEEE Trans. Pattern Analysis and Machine Intell.*, no. 6, pp. 679–698, 1986.