Circular Targets for 3D Alignment of Video and Lidar Sensors

Vincent Fremont, Sergio Alberto Rodriguez Florez, Philippe Bonnifait

Circular Targets for 3D Alignment of Video and Lidar Sensors

Vincent Fremont[1,2,*], Sergio A. Rodriguez F.[3,4], Philippe Bonnifait[1,2]

[1] *Université de Technologie de Compiègne*

[2] *CNRS Heudiasyc UMR 7253,60200 Compiègne, France*

[3] *Université Paris Sud, Orsay*

[4] *CNRS Institut d'Eléctronique Fondamentale UMR 8622, 91405 Orsay, France*

[*]Corresponding author, email : vincent.fremont@hds.utc.fr

Abstract

This paper presents a novel approach for solving the problem of 3D alignment between video and lidar sensors. The proposed method concerns intelligent vehicle applications, where the relative distance between sensor frames can be significant. Circular calibration targets are used in order to make full use of the perception properties of both lidar and video cameras, which greatly simplifies the calibration task. The method determines the relative pose in rotation and translation of the sensors, using sets of corresponding circular features acquired for several configurations of the targets. A performance analysis in simulation and an error propagation analysis are carried out. The calibration procedure is tested on different configurations, and the calibration accuracy and estimation of confidence intervals are evaluated on real data.

keywords: Extrinsic Calibration, Lidar, Vision Sensor, Circular targets, Intelligent Vehicles.

1 Introduction

For many intelligent vehicle applications, multi-sensor systems are a requirement in the development of complete perception architectures, such as when implementing Advanced Driver Assistance Systems or performing Autonomous Navigation Tasks. These multi-sensor systems usually comprise radars, lidars and video cameras, and the relative poses (rotation and translation) between the different sensors' coordinate systems have to be known. This becomes more of a problem when the relative distance between the sensors is significant, and when their perception fields are different (see Fig. 1).

One way of solving the extrinsic calibration problem is to use expensive systems such as theodolites together with specific markers corresponding to key positions of the sensors. In this case, the user has to be a calibration expert. Another way of addressing the problem is to implement an auto-calibration approach that harnesses knowledge of the movement of the mobile platform obtained, for example, via inertial measurement systems, to compensate for

Figure 1: Experimental Vehicle CARMEN with its 4-Layers lidar and Vision Systems

sensing delays and facilitate the association of natural features. Here no user intervention is needed, but the system can give poor results if the auto-calibration procedure is applied in suboptimal conditions. Using dedicated targets can be seen as falling midway between these two alternative strategies. The particular shape of the target is used to facilitate the matching of the features detected by the sensors in order to achieve a high-accuracy relative pose estimation. Dedicated targets are the choice we adopt here in order to have an easy-to-use and efficient calibration system. When calibrating a system comprising a lidar and a vision system, the data association cannot be done directly in the images, since the lidar produces near-infrared measurements that cannot be seen in images. We propose using circular targets for the correct alignment of video and lidar sensors. The calibration procedure uses several target poses and provides, via a batch process, the relative rotation and translation between the sensors with the associated confidence intervals.

Circular targets with holes have several advantages: they reduce significantly the lidar noise encountered when using the black-white zones of traditional "checkerboard" targets. Given that lidars are based on the reflection of diffuse light beams, black and white transitions reflect partial amounts of energy, which decreases the quality of the range measurements. It is much more accurate to perform a circle fitting in the 3D space of the lidar impacts contained within the perforation border of the calibration target. Finally, if the target has concentric circles, geometric and algebraic constraints can be harnessed to obtain an estimation of the camera-intrinsic parameters (i.e. focal length and principal point) simultaneously.

The paper is organized as follows. First, related works are summarized in Section 2. The 3D calibration procedure is then detailed in Section 3. Next, we present a comparison with other methods in Section 4. A performance analysis is carried out in simulation in Section 5. Real experiments are reported in Section 6 to illustrate the quality of results with different set-ups.

2 Related Works

Several calibration methods have been proposed recently for estimating 3D relative pose between lidars and video sensors. The authors of [1] propose a target-based calibration method between a camera and a single-row laser range finder. This method has been improved by [2]. A target-less calibration approach for a 3D lidar is presented in [3]. Similarly, [4] uses geometric primitives defined by the user in both perception modalities. In [5], the calibration procedure between an omnidirectional sensor and a lidar uses 3D planes and assumes that the

vision sensor has been previously calibrated. In [6] a Maximum Likelihood calibration procedure is proposed using a checkerboard observed at different poses for a calibrated camera and a multi-layer lidar. In [7], the authors describe a calibration procedure for a single-layer lidar and a camera mounted on it, based on a calibration target consisting of adjacent rectangular black and white boxes. In [8], the calibration makes use of relative pose data provided by an Inertial Measurement Unit (IMU) and therefore does not need specific targets to calibrate a system comprising 4 video feeds and a synchronous Velodyne lidar. A calibration procedure for lidar-camera rigs has been also proposed by [9]. The authors have developed a robust algorithm that uses correspondences between image line and 3D points. Finally, in [10], a 3D calibration pattern based on Y structures and pins of different lengths and colors is proposed. This particular shape provides information for the localization of corresponding pixels in the image.

3   Calibration Procedure

The problem is to estimate the relative pose , denoted $^{\mathcal{C}}[\mathbf{R}, \mathbf{t}]_{\mathcal{L}}$, between the lidar and a vision system using a calibration target with a circular hole, so it can be detected by both of these sensors (See Fig. 2). To this end the method estimates in each sensor frame the center and the normal to the supporting plane of the circle in space, representing the circular hole of the target.
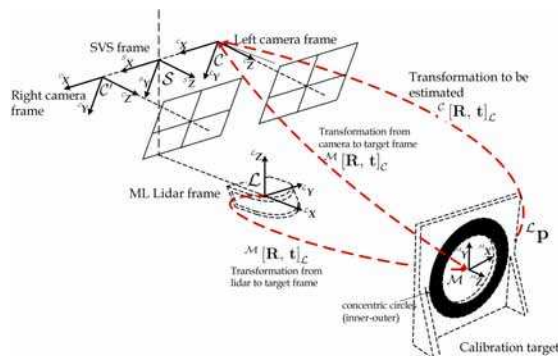

Figure 2: Frames involved in the lidar-camera calibration

3.1   Target Pose Estimation in the lidar Frame

The IBEO Alasca XT lidar considered in this study is equipped with a rotating mirror and four independent photo-diode receivers. This sensor is able to report 200m range measurements from four different detection planes, also called layers (see Fig. 3).

The four crossed scan layers cover a vertical field-of-view (FOV) of 3.2° where the inter-layer divergence corresponds to 0.8°. The advertised light beam divergence is 0.5 mrad, which is equivalent to a spot of diameter 50 cm at a range of 100 m. The multi-layer technology is intended to ensure object detection in the case of changes in vehicle pitch. This kind of lidar consequently has a
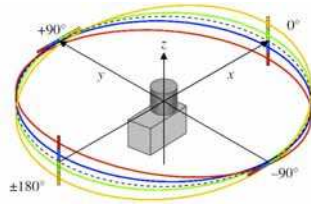
Figure 3: The lidar measures ranges in four detection planes (i.e. layers) with different directions

very limited FOV compared to a Velodyne sensor, which complicates the 3D alignment problem.

We shall assume that the intrinsic parameters of the lidar are known precisely: the 3D points that we use are assumed to be accurate; only measurement noise needs to be taken into account. For a 12.5 scanning frequency, the angular resolution is illustrated in Fig. 3, covering a total horizontal FOV of almost 330°.

Geometrically, a 3D laser impact can be defined in the lidar frame $\mathcal{L}$ as a single 3D point $^{\mathcal{L}}\mathbf{p}$. When intersecting the inner hole of the target, the lidar beam impacts will be contained within a 3D circle.

Before using the range information, a preprocessing stage is necessary to obtain a clean set of measurements using several lidar scans. Outlier data can then be filtered using the median filtering technique proposed in [2] and applied on n lidar scans of the calibration scene. Using the filtered data, points lying inside the perforation border of the calibration target are extracted using a 1-D edge detection: for each layer, a numerical derivative is performed on the depth value within a fixed range value where the target should be. The depth value signal of the target has a particular shape (see the left-hand side of Fig. 4) which gives 4 peaks when differentiated (see the right-hand side of Fig. 4). Taking only the two inner peaks of the derived depth value gives the perforation border of the calibration target. Since the lidar has 4 layers, $p = 8n$ points within the perforation border of the calibration target are extracted.
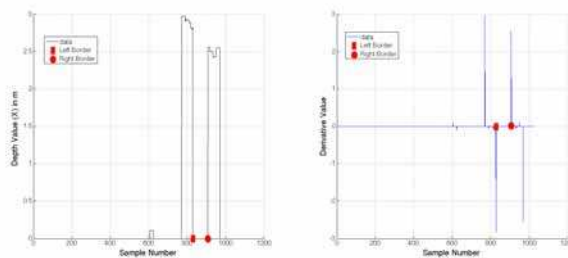

Figure 4: Calibration target perforation border detection using 1-D edge detection

Now, let $^{\mathcal{L}}\mathbf{c}$ be the center of the 3D circle, $^{\mathcal{L}}\mathbf{n}$ the normal vector to the supporting plane and r the known radius of the inner perforation border of the target.

As proposed in [11], a 3D circle fitting may be performed by estimating the

4

normal vector $^{\mathcal{L}}\mathbf{n}$ of the circle supporting plane and the circle center $^{\mathcal{L}}\mathbf{c}$, using a nonlinear minimization over the following geometrical criterion:

$$e \quad = \quad \sum_{i=1}^{p}[A_i^2 + B_i^2] \tag{1}$$

with

$$A_i =^{\mathcal{L}} \mathbf{n} \cdot \left(^{\mathcal{L}}\mathbf{p}_i -^{\mathcal{L}} \mathbf{c}\right) \text{ and } B_i = \|^{\mathcal{L}}\mathbf{n} \times \left(^{\mathcal{L}}\mathbf{p}_i -^{\mathcal{L}} \mathbf{c}\right)\|_2 - r \tag{2}$$

where the dot and cross product operators are denoted respectively by $(\cdot)$ and $(\times)$. The symbol $\|.\|_2$ denotes the L2-Norm, and

- A_i corresponds to the Euclidean distance between a lidar impact around the target hole , $^{\mathcal{L}}\mathbf{p}_i$, and the 3D plane defined by its normal vector $^{\mathcal{L}}\mathbf{n}$ and the circle center, $^{\mathcal{L}}\mathbf{c}$.

- B_i represents the Euclidean distance between a lidar impact around the target hole border, $^{\mathcal{L}}\mathbf{p}_i$, and the 3D circle axis passing through the circle center, $^{\mathcal{L}}\mathbf{c}$ in the direction $^{\mathcal{L}}\mathbf{n}$.

Accordingly, the criterion in Eq. 1 is minimized using the Levenberg-Marquardt algorithm (LM-algorithm) [12]. After convergence of the nonlinear minimization algorithm, applying this technique to various poses of the calibration target gives not only a first set of 3D laser features (i.e. circle centers, $^{\mathcal{L}}\mathbf{c}$, and normal plane vectors, $^{\mathcal{L}}\mathbf{n}$) but also a 3D circle reconstruction in the lidar frame for every pose.

3.2 Target Pose Estimation in the Camera Frame

In contrast to a lidar, a video camera is a passive sensor and its perception principle relies on the convergence of the visible light onto its imager. This study considers the image formation process through an ideal pinhole camera model. The strategy for estimating the target pose (i.e. its circle center and its normal vector) by the means of two imaged concentric circles is described below.

3.2.1 Geometrical Model of the Vision Sensor

Let $^{\mathcal{M}}\mathbf{p} = [X, Y, Z, 1]^T$ be a 3D homogeneous point in the target frame $\mathcal{M}$. This 3D point can be expressed in the camera frame $\mathcal{C}$ using the relative pose in rotation $^{\mathcal{C}}\mathbf{R}_{\mathcal{M}} = [\begin{array}{ccc} \mathbf{r_1} & \mathbf{r_2} & \mathbf{r_3} \end{array}]$ and translation $^{\mathcal{C}}\mathbf{t}_{\mathcal{M}} = [\begin{array}{ccc} t_x & t_y & t_z \end{array}]^T$:

$$^{\mathcal{C}}\mathbf{p} = \left[\begin{array}{cc} ^{\mathcal{C}}\mathbf{R}_{\mathcal{M}} & ^{\mathcal{C}}\mathbf{t}_{\mathcal{M}} \\ \mathbf{0} & 1 \end{array}\right] {}^{\mathcal{M}}\mathbf{p} \tag{3}$$

Let $\mathbf{x} = [u, v, 1]^T$ be the 2D homogeneous coordinates of its projection in the image of the left-hand camera frame $\mathcal{C}$ (see Fig. 2). Using the central projection model with homogeneous coordinates, we obtain the classical perspective camera projection equation [13]:

$$\mathbf{x} \sim \left[\begin{array}{ccc} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{array}\right] [\begin{array}{ccc} \mathbf{I} & | & \mathbf{0} \end{array}] {}^{\mathcal{C}}\mathbf{p} \tag{4}$$

where f represents the focal length, $(p_x, p_y)^T$ are the coordinates of the principal point and the notation $\sim$ to denote an equality up to a scale factor, since we are considering homogeneous coordinates.

The pinhole camera model assumes that the image coordinates are Euclidean coordinates having the same scales in the two axial directions. But image coordinates are measured in pixels, and therefore we need to consider unequal scale factors in the image axis. To this end the matrix $\mathbf{K}$ is modified in:

$$\mathbf{K} = \begin{bmatrix} f_x & f.s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} k_u & s & 0 \\ 0 & k_v & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix} \tag{5}$$

where the parameters f_x, f_y, u_0, v_0 and s are known as the intrinsic parameters and represent respectively the focal lengths in pixels in the x and y directions, the pixel coordinates of the principal point, and the skew, and k_u, k_v are the number of pixels per unit distance in image coordinates. Therefore the full perspective projection model is

$$\mathbf{x} \sim \mathbf{K} \begin{bmatrix} {}^{\mathcal{C}}\mathbf{R}_{\mathcal{M}} & | & {}^{\mathcal{C}}\mathbf{t}_{\mathcal{M}} \end{bmatrix} {}^{\mathcal{M}}\mathbf{p} \tag{6}$$

3.2.2 Vision Sensor Intrinsic Parameter Estimation

It should be remarked that prior to the intrinsic calibration, the lens distortions have to be removed. For this purpose, we use the method proposed by [14], hence the camera model presented in Eq. 6 is accurate. As mentioned earlier, in our approach, the circular calibration target also includes a second concentric circle around the circular hole. This is because projective properties of constrained conics make them suitable for camera calibration, as proven in [15, 16, 17]. It is therefore possible to infer metric properties about the camera (intrinsic parameters) and the scene (3D calibration target position) from the perspective projections of these circular features with minimal user interaction. The intrinsic calibration procedure, recalled in Appendix A, is applied to each camera before the extrinsic calibration step.

So far, all intrinsic parameters in Eq. 5 can be estimated (assuming skew $s = 0$). The next stage is to obtain the relative 3D transformation ${}^{\mathcal{C}}[\mathbf{R}, \mathbf{t}]_{\mathcal{M}}$ between the target and the vision sensors' frames.

3.2.3 Vision Sensor Pose Estimation

Let us now define a reference plane in frame $\mathcal{M}$ with equation ${}^{\mathcal{M}}Z = 0$. Thus, Eq. 6 becomes

$$\mathbf{x} \sim \mathbf{K} \begin{bmatrix} \mathbf{r_1} & \mathbf{r_2} & {}^{\mathcal{C}}\mathbf{t}_{\mathcal{M}} \end{bmatrix} . {}^{\mathcal{M}}\underline{\mathbf{p}} \sim \mathbf{H}.{}^{\mathcal{M}}\underline{\mathbf{p}} \tag{7}$$

where ${}^{\mathcal{M}}\underline{\mathbf{p}} = [X, Y, 1]^T$. The equation of a 3D circle of radius r on the reference plane and centered at the origin of the frame $\mathcal{M}$ can be defined as:

$$ {}^{\mathcal{M}}\underline{\mathbf{p}}^T . \mathbf{Q} . {}^{\mathcal{M}}\underline{\mathbf{p}} = 0, \quad \mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -r^2 \end{bmatrix} $$

Using Eq. 7, the perspective projection of the circle $\mathbf{Q}$ gives on the image plane a conic $\mathbf{C}$ which is defined by

$$\mathbf{x}^T \mathbf{C} \mathbf{x} = 0, \quad \mathbf{C} \sim \mathbf{H}^{-T} \mathbf{Q} \mathbf{H}^{-1} \tag{8}$$

The conic C is estimated from the segmented image [17] using a classical ellipse fitting algorithm [18]. Now, let $\tilde{\mathbf{C}}$ be a generic proper conic (non-singular with 2 positive and 1 negative eigenvalues):

$$\tilde{\mathbf{C}} = \mathbf{K}^T \mathbf{C} \mathbf{K} \tag{9}$$

The homogenous image coordinates of the projected circle center $\mathbf{x_c} = (x_c, y_c, 1)^T$ are obtained by projecting the frame origin, which is the center of $\mathbf{Q}$, in the image:

$$\mathbf{x_c} = (x_c, y_c, 1)^T \sim \mathbf{K} \begin{bmatrix} {}^{\mathcal{C}}\mathbf{R}_{\mathcal{M}} & | & {}^{\mathcal{C}}\mathbf{t}_{\mathcal{M}} \end{bmatrix} (0, 0, 0, 1)^T \sim \mathbf{K}^{\mathcal{C}}\mathbf{t}_{\mathcal{M}} \tag{10}$$

Next, the center of $\mathbf{Q}$ is by definition the pole of the line at infinity [19], hence in the image, the projected center of $\mathbf{Q}$ is the pole of the vanishing line:

$$\mathbf{l}_{\infty} \sim \mathbf{C} \mathbf{x_c} \tag{11}$$

then

$$\begin{aligned} \mathbf{C} \mathbf{x_c} &\sim & \mathbf{C} \mathbf{H} (0, 0, 1)^T \\ &\sim & \mathbf{H}^{-T} \mathbf{Q} \mathbf{H}^{-1} \mathbf{H} (0, 0, 1)^T = -r^2 \mathbf{H}^{-T} (0, 0, 1)^T \\ &\sim & \mathbf{r_3} = {}^{\mathcal{C}}\mathbf{n} \end{aligned} \tag{12}$$

Therefore, the normal vector to the target plane containing the circle $\mathbf{Q}$ is given by:

$$^{\mathcal{C}}\mathbf{n} \sim \tilde{\mathbf{C}} \mathbf{K}^{-1} (x_c, y_c, 1)^T \tag{13}$$

As presented in [15], it is possible to obtain the projected center of two concentric circles using only their imaged conics. Basically, the projected circle center of two concentric circles is different from the centers of the projected conics, but always lies on a line defined by them. It is then possible to use the cross-ratio of four aligned points to extract the image coordinates of the projected circle center $\mathbf{x_c}$.

Thus far we have obtained only the orientation of the target in the camera frame. The last unknown is the coordinates of center of the 3D circle in the camera frame. Let us consider the line containing the center of the camera frame $^{\mathcal{C}}\mathbf{c_c} = (0, 0, 0)^T$ and the canonical coordinates of the projected circle center $^{\mathcal{C}}\mathbf{x_c} = \mathbf{K}^{-1} (x_c, y_c, 1)^T$ in 3D space. Points on this 3D line are represented by

$$\mathbf{x}(\lambda) = {}^{\mathcal{C}}\mathbf{c_c} + \lambda \left({}^{\mathcal{C}}\mathbf{x_c} - {}^{\mathcal{C}} \mathbf{c_c} \right) = \lambda \mathbf{K}^{-1} (x_c, y_c, 1)^T \tag{14}$$

In the camera frame, since the 3D circle $\mathbf{Q}$ is centered at the origin of the frame $\mathcal{M}$, $^{\mathcal{C}}\mathbf{c} = {}^{\mathcal{C}}\mathbf{t}_{\mathcal{M}}$. Therefore we look for the value of λ which gives the position of $^{\mathcal{C}}\mathbf{c}$:

$$\begin{aligned} ^{\mathcal{C}}\mathbf{t}_{\mathcal{M}} &= & \lambda \mathbf{K}^{-1} (x_c, y_c, 1)^T \\ \Longleftrightarrow^{\mathcal{C}} \mathbf{n}^T {}^{\mathcal{C}}\mathbf{t}_{\mathcal{M}} &= & \lambda^{\mathcal{C}}\mathbf{n}^T \mathbf{K}^{-1} (x_c, y_c, 1)^T \\ \Rightarrow \lambda &= & \frac{^{\mathcal{C}}\mathbf{n}^T {}^{\mathcal{C}}\mathbf{t}_{\mathcal{M}}}{^{\mathcal{C}}\mathbf{n}^T \mathbf{K}^{-1} (x_c, y_c, 1)^T} \\ \Rightarrow^{\mathcal{C}} \mathbf{c} &= & \frac{^{\mathcal{C}}\mathbf{n}^T {}^{\mathcal{C}}\mathbf{t}_{\mathcal{M}} \mathbf{K}^{-1} (x_c, y_c, 1)^T}{^{\mathcal{C}}\mathbf{n}^T \mathbf{K}^{-1} (x_c, y_c, 1)^T} \end{aligned} \tag{15}$$

7

The value $d =^{\mathcal{C}} \mathbf{n}^T \,^{\mathcal{C}}\mathbf{t}_{\mathcal{M}}$ is the orthogonal distance to the supporting plane of the circle. From [20] we have:

$$d = \sqrt{\rho^3}\, r \tag{16}$$

where ρ is the smallest positive eigenvalue of $\tilde{\mathbf{C}}$ (see Eq. 9) with $\det\left(\tilde{\mathbf{C}}\right) = -1$ and r is the radius of the corresponding circle $\mathbf{Q}$:

$$^{\mathcal{C}}\mathbf{c} = \frac{d\,\mathbf{K}^{-1}\,(x_c,\,y_c,\,1)^T}{^{\mathcal{C}}\mathbf{n}^T\,\mathbf{K}^{-1}\,(x_c,\,y_c,\,1)^T} \tag{17}$$

Like in Section 3.1 for the lidar, we now have the orientation $^{\mathcal{C}}\mathbf{n}$ and the position $^{\mathcal{C}}\mathbf{c}$ of the circular target in the video camera frame. The next step, presented in Section 3.3, is to match this information and to calculate the rotation and translation between the lidar and the vision sensors.

3.3 Lidar to Camera Transformation Estimation

The lidar-to-camera calibration procedure works as a two-step process: firstly, the transformation parameters are calculated in closed form using the sensor-to-target calibration. Secondly, the transformation is refined by minimizing a point-to-point error measure, similar to the Iterative Closest Point Algorithm.

3.3.1 Initialization of the procedure

A well-known closed-form solution for this problem is the method developed by [21]. This method consists in obtaining the optimal rotation from the Singular Value Decomposition (SVD) of the correlation matrix of the centered point sets represented by $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \left[^{\mathcal{L}}\hat{\mathbf{c}}_i -^{\mathcal{L}} \overline{\mathbf{c}}\right]_{3\times n} \left[^{\mathcal{C}}\hat{\mathbf{c}}_i -^{\mathcal{C}} \overline{\mathbf{c}}\right]_{n\times 3}^T = \mathbf{U}\mathbf{S}\mathbf{V}^T \tag{18}$$

where n is the number of poses, and $^{\mathcal{L}}\hat{\mathbf{c}}_i$ are the estimated coordinates of the 3D circle center from the ith pose of lidar measurements. $^{\mathcal{L}}\overline{\mathbf{c}}$ is the centroid of the 3D-circle center point set in the lidar frame, $^{\mathcal{C}}\hat{\mathbf{c}}_i$ are the coordinates of the 3D-circle center point set estimated from the ith pose of camera measurements and $^{\mathcal{C}}\overline{\mathbf{c}}$ is the centroid of the 3D-circle center point set in the camera frame. Therefore, the 3×3 optimal rotation matrix is obtained as follows:

$$^{\mathcal{C}}\mathbf{R}_{0,\mathcal{L}} = \mathbf{V}\mathbf{U}^T \tag{19}$$

where the subscript (0) denotes that it constitutes an initial estimation. The translation vector $^{\mathcal{C}}\mathbf{t}_{0,\mathcal{L}}$ is obtained as the vector which aligns the centroid of the 3D-circle center point set in the camera frame, $^{\mathcal{C}}\overline{\mathbf{c}}$, and the rotated centroid $^{\mathcal{C}}\mathbf{R}_{0,\mathcal{L}}.^{\mathcal{L}}\overline{\mathbf{c}}$:

$$^{\mathcal{C}}\mathbf{t}_{0,\mathcal{L}} =^{\mathcal{C}} \overline{\mathbf{c}} -^{\mathcal{C}} \mathbf{R}_{0,\mathcal{L}}.^{\mathcal{L}}\overline{\mathbf{c}} \tag{20}$$

3.3.2 Non-linear optimization of the solution

This solution $^{\mathcal{C}}[\mathbf{R}_0, \mathbf{t}_0]_{\mathcal{L}}$ is usually a good starting guess for the extrinsic calibration. Therefore, with the aim of refining these estimated parameters, the first step is to generate the 3D circles of the n poses estimated by the camera. This consists in computing m points of every estimated circle pose by using the 3D circle center and an orthonormal base lying within the plane of the circle. The orthonormal base is obtained from the normal vector to the plane of the circle, by applying the Gram-Schmidt procedure [22]. Let $^{\mathcal{C}}\mathbf{p}_{i,k}$ be the kth generated 3D point using the camera estimation of the ith pose. The second step is to generate the 3D circles of all the poses estimated by the lidar (Section 3.1) just like for the camera estimations, giving $^{\mathcal{L}}\mathbf{p}_{i,k}$. The first guess is then systematically applied for the rigid transformation $^{\mathcal{C}}[\mathbf{R}_0, \mathbf{t}_0]_{\mathcal{L}}$ so as to obtain the points in the camera frame. The third step, assuming that the error orientation of the first guess rigid transformation is lower than $\pi/2$, is to match the 3D points of the camera and lidar transformed estimations for every pose using the nearest neighbor criterion based on the Euclidean distance:

$$min_{k,j} \left\| {}^{\mathcal{C}}\mathbf{p}_{i,k} - ({}^{\mathcal{C}}\mathbf{R}_{0,\mathcal{L}}.{}^{\mathcal{L}}\mathbf{p}_{i,j} + {}^{\mathcal{C}}\mathbf{t}_{0,\mathcal{L}}) \right\|_2 \tag{21}$$

An example of 3D point matching is presented in Fig. 5. At this point it is worth mentioning that we have a corresponding 3D point set of camera and lidar observations. Finally, the refining of the rigid transformation parameters, $[\alpha, \beta, \gamma, t_x, t_y, t_z]^T$, is obtained by minimizing the following non-linear objective function:

$$\epsilon = \sum_{i=1}^{n} \sum_{k=1}^{m} \mathbf{W}.D_{ik}^2 \tag{22}$$

with

$$D_{ik} = \left\| {}^{\mathcal{C}}\mathbf{p}_{i,k} - {}^{\mathcal{C}}\mathbf{R}_{\mathcal{L},(\alpha, \beta, \gamma)}.{}^{\mathcal{L}}\mathbf{p}_{i,k} - {}^{\mathcal{C}}\mathbf{t}_{\mathcal{L},(t_x, t_y, t_z)} \right\|_2 \tag{23}$$

where D_{ik} represents the Euclidean distance residual of the points after applying the rigid transformation and $\mathbf{W}$ is a weighting matrix. The results are obtained using a robust M-estimator algorithm for calculating the robust weights as in [23] and the LM-algorithm. After convergence, the solution of the calibration problem is represented by $^{\mathcal{C}}[\mathbf{R}, \mathbf{t}]_{\mathcal{L}}$.
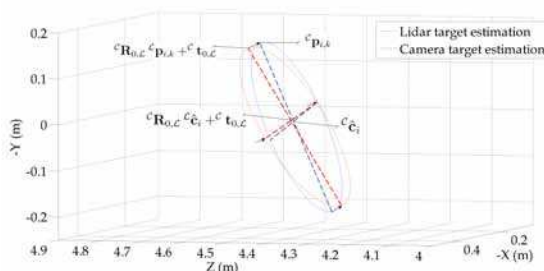

Figure 5: Matching camera and lidar target pose estimations

3.4 Calibration Confidence Estimation

The precision of the calibration results can be estimated under the assumption that measurement errors are normally distributed. Therefore, the covariance matrix of the estimated parameters, $\mathbf{C}_\sigma$, is defined as follows:

$$\mathbf{C}_\sigma = \sigma^2 . \left(\mathbf{J}^T \mathbf{J}\right)^{-1} \tag{24}$$

with $\sigma^2 = \frac{1}{\theta - \varphi} \|\epsilon\|^2$. $\mathbf{J}$ represents the Jacobian matrix of the last LM-algorithm iteration and σ^2 represents an unbiased estimate of the variance, defined by θ, the number of observations, φ, the number of estimated parameters and ϵ the residual of the non-linear objective function (see Eq. 22). In our case, φ is equal to 6 (3 rotations and 3 translations) and $\theta - \varphi$ represents the degrees of freedom of the Student's t-distribution. Based on the above classical approach for the covariance matrix of the non-linear fitted parameters [24], the width of the $100(1 - \tau)\%$ confidence interval is given by:

$$\delta \mathbf{C}_i = t_{\theta - \varphi}^{\tau/2} . \sqrt{\mathbf{C}_\sigma \left(i,\, i\right)} \tag{25}$$

where $\sqrt{\mathbf{C}_\sigma \left(i,\, i\right)}$ is the standard deviation of the ith estimated parameter and $t_{\theta - \varphi}^{\tau/2}$ the corresponding t-value for τ and $\theta - \varphi$ degrees of freedom.

Matlab users may easily obtain the confidence intervals for the parameters via the NLPARCI function, and both the residuals and the Jacobian matrix via the LSQNONLIN function, which performs the non-linear minimization.

3.5 Calibration Algorithm

The overall approach can be summarized in the following algorithm:

Algorithm 3.1 Circle-based lidar-Vision sensors Calibration Algorithm
---
Input: - Images of the circular pattern (for at least $n = 6$ poses)
       - 3D points of the circular pattern from lidar (for at least 6 poses)
Output: $^{\mathcal{C}}\left[\mathbf{R},\, \mathbf{t}\right]_{\mathcal{L}}$

1: **for** $i = 1$ **to** n **do**
2:     ▶ Estimate the ith lidar calibration pose, $\left[^{\mathcal{L}}\mathbf{n},\, ^{\mathcal{L}}\hat{\mathbf{c}}\right]_i$, as described in Section 3.1.
3:     ▶ Estimate the ith camera calibration pose, $\left[^{\mathcal{C}}\mathbf{n},\, ^{\mathcal{C}}\hat{\mathbf{c}}\right]_i$, as described in Section 3.2.
4: **end for**
5: ▶ Compute a first guess, $^{\mathcal{C}}\left[\mathbf{R}_0,\, \mathbf{t}_0\right]_{\mathcal{L}}$, for the lidar-camera rigid transformation using the linear solution (Section 3.3)
6: **repeat**
7:     ▶ Non-linear minimization using LM-algorithm according to the criterion Eq. 22
8:     ▶ Robust noise variance estimation σ^2 based in non-linear minimization residuals
9:     ▶ Weighting matrix $\mathbf{W}$ update
10: **until** convergence of $^{\mathcal{C}}\left[\mathbf{R},\, \mathbf{t}\right]_{\mathcal{L}}$

---

4 Comparison with other approaches

Table 1 presents a synthesis of different camera/lidar calibration procedures, which fall into two main categories: methods using user-selected corresponding features [3][4], and methods with a calibration pattern [1][2][6][7][9][10]. The main drawback with the first category is that lidars emit infrared light that cannot be seen by video cameras. Therefore corresponding features, like points or edges, have to be selected or detected and then precisely associated with

	Calibration Pattern	Camera-lidar Interdistance	Application	Error analysis	lidar type	Intrinsic calib.
Proposed	Perforated circular target	Important	Intelligent Vehicle (IV)	yes	4-layers	yes
[1][2]	Chessboard Planar target	small	IV	yes	1-layer	yes
[3]	Hand feature	small	IV	yes	1-layer	no
[4]	Hand feature	Important	IV	no	1-layer	no
[6]	Chessboard Planar target	Important	IV	no	4-layers	yes
[7]	3D boxes	small	robotic, IV	no	1-layer	yes
[8]	IMU Pose data	Important	IV	yes	multi-layers	yes
[9]	Planar target	small	robotic	yes	1-layer	no
[10]	3D calib. pattern	small	robotic	no	1-layer	no

Table 1: Comparison of lidar-to-camera 3D calibration methods (IV : Intelligent Vehicle)

their image projection, over several poses. In other words, a data association problem has to be solved to ensure that it is sufficient to find the relative position between the sensors. The second category also poses a particular problem for the calibration: the pattern has to be chosen so that it can be seen by both modalities (lidar and vision). So far, several works [1][2][9] have achieved this using a planar calibration pattern for both single and multi-layer lidar. It is easy to provide a precise estimation of the camera parameters using this kind of calibration target, but it is difficult to get a common reference point from the lidar data, because only the normal and the orthogonal distance can be obtained. Therefore additional constraints need to be used to find a singularity-free solution. [7][10] use a 3D calibration target that can be perceived and easily matched via both sensing capabilities. In our work, the hole in the calibration target ensures an easy-to-detect 3D target without singularities. [8] gives an alternative way of calibrating a lidar-camera system using relative pose data from an external sensing system (here an IMU[1]). The proposed algorithm opens the way to fully automatic auto-calibration, and appears very promising, even if less accuracy is achieved.

5   Performance Analysis in Simulation

Four simulation trials are reported below, where the accuracy and consistency of the proposed calibration routine are respectively quantified and verified. In order to have conditions similar to the real multi-sensor system, the simulation model corresponds to the sensor relative position on board the vehicle.

---

[1]Inertial Measurement Unit

The extrinsic parameters are the translation vector ${}^{\mathcal{C}}\mathbf{t}_{\mathcal{L}} = [-0.2,\, 0.8,\, 1.8]^T$ in meters and an orientation matrix ${}^{\mathcal{C}}\mathbf{R}_{\mathcal{L},(\alpha,\beta,\gamma)}$, computed from the rotation angles $\alpha = 11°$, $\beta = -1°$ and $\gamma = 0.5°$ respectively in the X, Y and Z axes.

The lidar impacts were generated as the intersection of the lidar beam and the simulated calibration target and all lidar layers intersected the calibration object. A Gaussian white noise was added in the direction of the lidar beam, simulating range measurement errors. Based on the model given in Eq. 6, a synthetic image of the calibration target was generated. A Gaussian white noise was added to the projected coordinates of the calibration target and to the intrinsic parameters.

5.1 Test No. 1: Influence of the number of poses

A first Monte Carlo-like simulation was performed in order to estimate the precision achieved by the proposed method using a minimal number of poses (worst case). To this end, six random poses were distributed and oriented randomly in the common field-of-view of the multi-sensor system over 100 trials. The noise added to the image coordinates and the focal lengths was fixed to one pixel. Here, the intrinsic camera parameters were constrained to a unitary aspect ratio (i.e. $f_x = f_y = 1670$) and a principal point at the image center.

At each trial, the extrinsic parameters were estimated. The results give a relative position error of 46.1mm and a relative orientation error of 3.4 degrees. Six poses would appear not to be accurate enough, considering that the obtained calibration results lead to unacceptable image projection errors for objects located at the advertised range of the lidar (200m). A complementary test was therefore performed to observe the error as the number of poses changes. As shown in Fig. 6, an improvement is obtained as the number of poses increases. Seven to nine poses for calibration gives a good trade-off between precision and the overhead of pose acquisition.
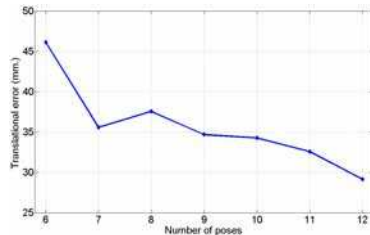

Figure 6: Extrinsic calibration parameter error behavior by poses

5.2 Test No. 2: Influence of image noise

In order for the error behavior of the method to be evaluated with respect to an image noise variation, a second simulation test was performed. Similarly to the first simulation test, one hundred trials using seven poses were generated for each level of the image noise. The Gaussian white noise added between one and three pixels of standard deviation. So as to compare the results obtained by the robust non-linear minimization of the 3D poses, we also executed an Iterative Closest Point algorithm , a reference among classical registrations of

3D point sets. The results given in Fig. 7 demonstrate that the proposed robust registration scheme performs quite well and much better than the classical one.
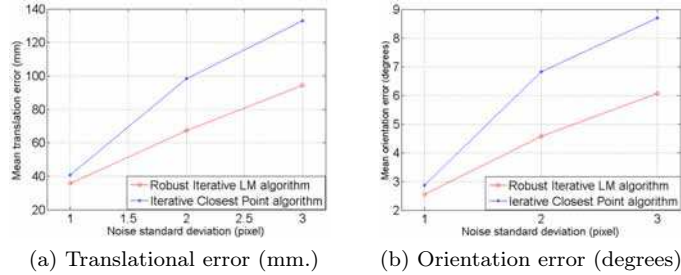

(a) Translational error (mm.)   (b) Orientation error (degrees)

Figure 7: Extrinsic calibration errors for different image noise levels

5.3 Test No. 3: Consistency test

The consistency of the calibration estimation, presented in Section 3.4, is crucial for error propagation and uncertainty quantification when merging lidar and vision data. Since the estimated intervals correspond to a given probability (e.g. 95% confidence) of the solution lying between them, a consistency test was performed. To this end, the estimations of the rigid transformations and the confidence intervals using 7 poses in the calibration process were plotted over a hundred trials.

Fig. 8 illustrates, as an example, the consistency test for the orientation angle parameter in the y-axis of the lidar frame (i.e. pitch vehicle angle) previously denoted β. Results are also compared with those obtained through the Iterative Closest Point registration. In the figure, the estimates are centered so that the ground truth value is the x plot axis.
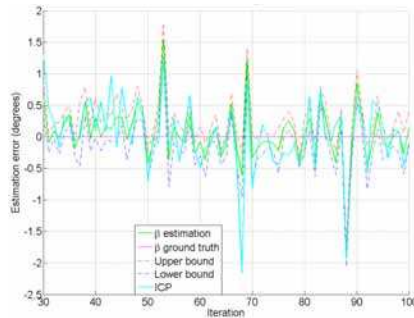

Figure 8: Consistency test for β at 95%

It will be remarked that the robust non-linear algorithm runs significantly better than the Iterative Closest Point registration. The two approaches nevertheless oscillate in the same way, since the same criterion, based on the L2-norm distance between two point-sets, is minimized. The main difference compared to the ICP approach is that we know the point-to-point correspondence (see Section 3.3.2). Looking at Fig. 8, it will also be remarked that the consistency of the estimates is slightly optimistic, since several intervals do not contain the

zero value. Moreover, it is worth mentioning that the convergence is globally very good, even if it is not guaranteed, as exemplified by Fig. 8 at the 53rd, 69th and 88th trials.

5.4   Test No. 4: Influence of the projection camera model

Focusing now on the behavior of the extrinsic calibration method with regard to the projection camera model parameters, a new Monte Carlo-like test was performed with fifty trials each. In this test, seven random poses were used and a Gaussian-white noise, ranging between one and three pixels of standard deviation, was added to the image coordinates. The projection camera model used during the calibration process was constrained to a unitary aspect ratio (i.e. $f_x = f_y$), in contrast to the simulated model.

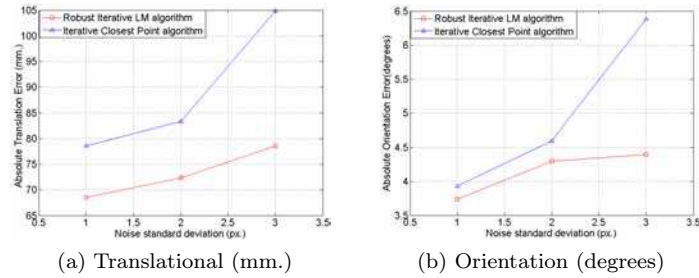(a) Translational (mm.)   (b) Orientation (degrees)

Figure 9: Absolute errors using a camera model with a unitary aspect ratio

For every trial, the extrinsic parameters were estimated and the mean error of the fifty trials was computed. The results obtained by the robust non-linear minimization of the 3D poses are presented in Fig. 9. It will be remarked that even though the parameter calibration error is not negligible, this method provides much better results than those obtained using the classical registration method.

A second test was performed taking in the same conditions, except that the camera model was unconstrained to a non-unitary aspect ratio. This test quantifies the influence of the camera model over the absolute error of the extrinsic calibration parameters. Fig. 10 illustrates the significant improvement obtained when using an unconstrained camera model.

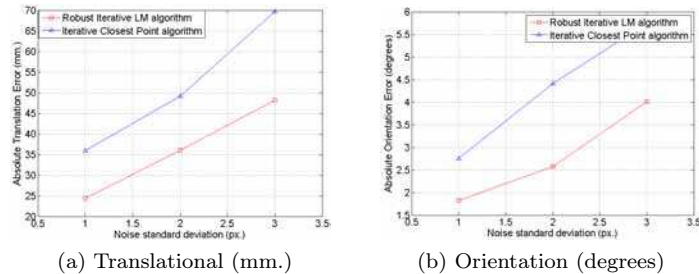(a) Translational (mm.)   (b) Orientation (degrees)

Figure 10: Absolute errors using an unconstrained camera model

The results of this last test reveal that the intrinsic parameters have a considerable influence over the extrinsic parameters. It is worth recalling that the extrinsic calibration method is based only on the minimization of the Euclidean error between the lidar and the camera perception. Therefore, errors in the intrinsic parameters will be compensated for by the extrinsic parameters, thus reducing this bias.

6   Real Experiments

The evaluation of the calibration method in real conditions was performed using the 4-layer lidar and different vision system configurations, namely monocular with short and wide angles, and a stereo vision rig. The experimental conditions are described below, and the results for both considered configurations are reported. The Matlab calibration source files are publicly available here[2].

6.1   Hardware setup

The CARMEN experimental vehicle (see Fig. 1) is used for implementing and testing perception functions in real-life conditions. This platform uses a number of sensors including the 4-layer lidar and the vision systems. The lidar is installed at the front of the vehicle in the bumper section, while the cameras are located behind the windshield and on the roof of the vehicle (see Fig.1). This configuration of sensors involves only occasional occlusions to the camera for short distances with respect to the lidar.

6.2   Alignment between the lidar and video camera

In this experiment an IBEO Alasca XT and a Sony DFW-VL500 camera were used (see Fig. 11). The resolution of the camera was set to 640×480 pixels. We used a calibration target with two concentric circles of radii 33cm and 23cm. The camera's focal distances f_x and f_y, and the principal point were estimated using the circle-based calibration target.

Twenty scans were performed for each pose in the calibration process. Only 7 poses were used to estimate the initial guess solution for the rigid transformation.

Figure 11: CARMEN experimental platform, showing the relative locations of the sensors

For the case of a unitary aspect ratio in the projection camera model, the extrinsic calibration parameters $^C[\mathbf{R}, \mathbf{t}]_{\mathcal{L}}$ were first estimated. Since the lidar

---

[2]http://www.hds.utc.fr/~vfremont/dokuwiki/doku.php?id=en:links

produces near infrared data, we use this rigid transformation to express these data in the camera frame, and a projection of the multi-layer range measurements onto the scene image can be obtained using the perspective projection model of Eq. 6. The results in the left-hand image are presented in Fig. 12.

In a second trial, we used an unconstrained aspect ratio camera model and recomputed the corresponding extrinsic parameters.

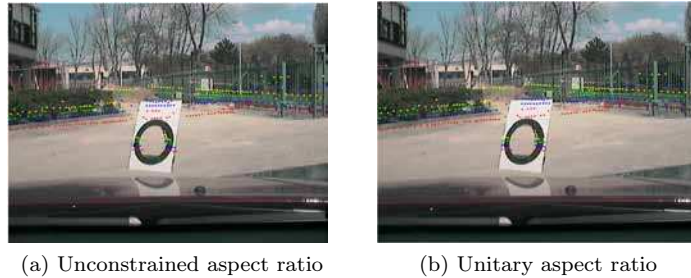(a) Unconstrained aspect ratio   (b) Unitary aspect ratio

Figure 12: Projection image of lidar data using the extrinsic calibration method

The image projection of the lidar data reveals that the errors resulting from the assumption of a unitary aspect ratio in the projective camera model are almost imperceptible. In Table 2, however, one observes that this simplifying assumption can provide estimates of the relative position of the sensors, which can be slightly biased (see the value of t_x in column 4). This phenomenon was also visible in the previous simulations.

The confidence intervals for the rotation angles are quite similar, which is not the case for the translation parameters. For instance, the confidence on t_x is twice as good as that on t_y, which is due to the aspect ratio of the CCD sensor. The method provides very high confidence intervals for the depth direction (i.e. t_z), which is only to be expected, since the geometrical configuration of the calibration pattern is optimal in this direction. Notice that the "measured values" in Table 2 are not ground truth, but were measured manually and are shown for comparison.

Table 2: Calibration results obtained with real data

Results of the Test using Real Data					
Translation (m)	Two Focal Camera	Confidence Interval	One Focal Camera	Confidence Interval	Measured
t_x	-0.1651	±0.0491	-0.3331	±0.0647	-0.2
t_y	0.9208	±0.1176	0.9246	±0.0648	0.88
t_z	1.8466	±0.0116	1.8027	±0.0087	1.82
Rot. angles (rad)					
$R_x = \alpha$	1.5370	±0.0296	1.5428	±0.0156	n.a.
$R_y = \beta$	-0.0455	±0.0118	-0.0505	±0.0149	n.a.
$R_z = \gamma$	1.6075	±0.0466	1.6296	±0.0261	n.a.

A second experiment was carried out using the IBEO Alasca XT and a

PointGrey camera with a Theia wide-angle lens (of focal length approx. 1.3mm). In Fig. 13, the detected 3D circle is visible on the left-hand side, corresponding to the camera image presented on the right-hand side. Notice that the center position and the orientation of the 3D circle are well estimated.

(a) Detected 3D circle using lidar data   (b) Corresponding wide-angle image
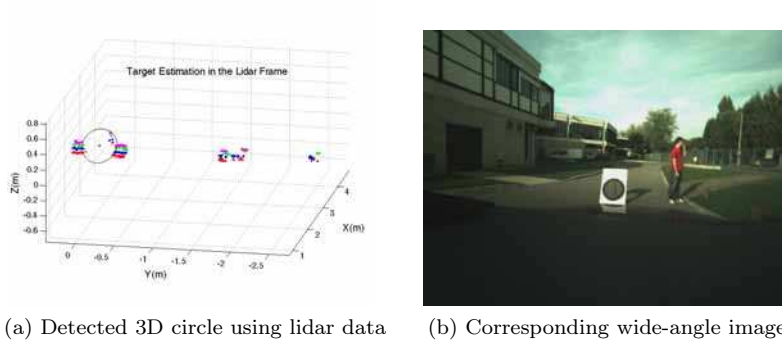
Figure 13: 3D circle detection

Fig. 14 shows the image re-projection of lidar data for two different poses, corresponding to the estimated calibration values presented in Tab. 3.
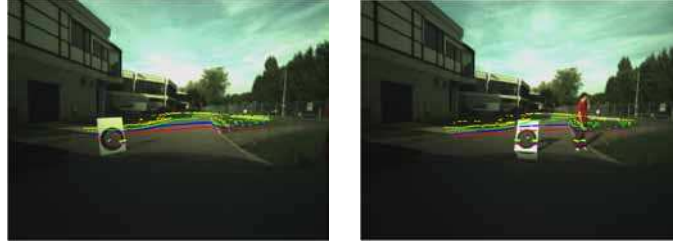

Figure 14: Lidar data re-projection onto two different images of the calibration video.

Table 3: Calibration results obtained with real data for 12 target poses

Results of the Test using Real Data			
Translation (m)	Wide Angle Camera	Confidence Interval	Measured
t_x	-0.053	±0.072	-0.1
t_y	0.954	±0.15	0.922
t_z	1.81	±0.05	1.87
Rot. angles (rad)			
$R_x = \alpha$	-0.036	±0.093	-0.035
$R_y = \beta$	-0.011	±0.025	0.047
$R_z = \gamma$	-0.0024	±0.03	0

This experiment gives good results for an unusual setup: the large depth between the sensors, the short focal length of the optical lens and the small size

Table 4: Results of stereo calibration (extrinsic parameters w.r.t left camera frame)

	Manufacturer data	Tsai for Left/Right Camera	Proposed for Left/Right Cameras
f_x(pix.)	216	223/220	230/226
f_y(pix.)	216	221/219	229/227
u_0 (pix.)	160	149/151	140/146
v_0 (pix.)	120	116/110	119/113
Rotation (rad)	$^{\mathcal{C}}[\alpha,\beta,\gamma]^T_{\mathcal{C}'} =$ $[0,0,0]^T$	$^{\mathcal{C}}[\alpha,\beta,\gamma]^T_{\mathcal{C}'} =$ $[-0.01,-0.007,-0.014]^T$	$^{\mathcal{C}}[\alpha,\beta,\gamma]^T_{\mathcal{C}'} =$ $[0.026,0.003,-0.016]^T$
Translation (mm)	$^{\mathcal{C}}[t_x,t_y,t_z]^T_{\mathcal{C}'} =$ $[-450,0,0]^T$	$^{\mathcal{C}}[t_x,t_y,t_z]^T_{\mathcal{C}'} =$ $[-442.6,-0.15,2.8]^T$	$^{\mathcal{C}}[t_x,t_y,t_z]^T_{\mathcal{C}'} =$ $[-447.7,6.2,5.5]^T$

of the target make the calibration estimation process difficult. Nevertheless, the results are coherent with respect to the measured values, and the re-projected lidar data are well placed in the images.

6.3 Calibration between a lidar and a stereo-vision system

Fifteen stereo images of the circular target were acquired at a low resolution of 320x240 pixels. Table 4 shows the calibration results for both intrinsic and extrinsic parameters of the stereo rig. Results are compared with the classical Tsai approach [25] and with the information provided by the camera manufacturer. Regarding the estimation of extrinsic parameters, our approach uses the target center coordinates $^{\mathcal{C}}\mathbf{c}_i$ and $^{\mathcal{C}'}\mathbf{c}_i$ calculated using Eq. 17 in both cameras $\mathcal{C}$ and $\mathcal{C}'$. Since each image pair i provides two 3D points, the 3D transformation between the two cameras can be estimated using the 3D registration approach [21] presented in Section 3.3.1. It can be seen (e.g. Table 4) that our algorithm gives similar results to Tsai's method, but our calibration target can also be used with a multi-layer lidar, since its shape can easily be detected in 3D.

In this calibration trial, eight different target poses were acquired. Each pose was composed of twenty lidar scans and the corresponding target image of the left-hand camera in the stereo rig. The rigid transformation between the lidar and the left camera was then computed and the intervals of confidence were estimated. The obtained relative orientation and position of the lidar in the camera frame are presented as an Euler and a translation vectors:

$$[\alpha\,\beta\,\gamma]^T = [\begin{array}{ccc} 0.09 \pm 0.031 & -0.003 \pm 0.013 & -0.037 \pm 0.012 \end{array}]^T \text{(in m.)}$$

$$[t_x\,t_y\,t_z]^T = [\begin{array}{ccc} 0.266 \pm 0.034 & 1.323 \pm 0.107 & 2.44 \pm 0.015 \end{array}]^T \text{(in rad.)}$$

It will be remarked that the z component of the translation vector is quite large (around 2.5m) as visible in Fig. 1. In order to project lidar scan points onto the right-hand image (see Fig. 15), the point coordinates from the left-hand camera frame need to be transformed into right-hand camera coordinates using the extrinsic parameters of the stereo rig given in Table 4. We wish to emphasize that in the case of stereo video cameras, our circular target means

that the 3D alignment between the lidar and the two cameras and the extrinsic calibration of the stereo rig can both be performed.
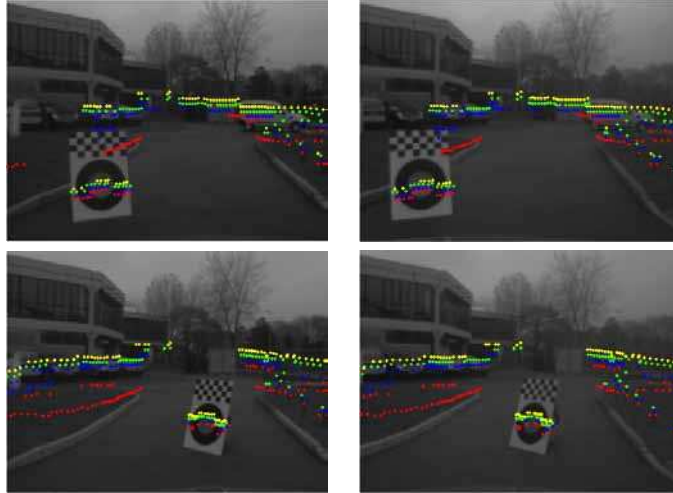

Figure 15: lidar data re-projection onto stereo images for two different target positions

7   Conclusion

In this paper, we proposed a new extrinsic calibration method for a configuration comprising 4-layer lidar and video cameras to be used in automotive applications. This kind of calibrated transformation is useful for multi-sensor obstacle detection, as shown in [26], where re-projected lidar information is employed to perform a visual confirmation of the detected objects.

Through the use of a circular calibration target, extrinsic calibration and intrinsic camera calibration can be performed simultaneously. Both computer simulations and real data were used to test the proposed technique, and confident results have been obtained. The estimation of the confidence intervals provided by the calibration method are useful for uncertainty propagation in data sensor fusion methods. It important to note that our method quite easily gives a good estimation of the relative pose of the sensors. One additional perspective of this work is providing confidence intervals corresponding to the extrinsic calibration parameters for the stereo rig, using only circular calibration targets. Another perspective is the estimation of the distortion parameters using concentric circles. We are also investigating how to analyze parameter sensitivity with respect to calibration target space positions, in order to provide more accurate results in both intrinsic and extrinsic parameters. Finally, we think that this method could easily be adapted to the calibration of a 3D lidar (e.g. Velodyne sensor) with respect to multi-camera systems. Further research might also be carried out to find the optimal size of the target, considering features of both lidar and video sensors, in particular with the use of wide-angle optical lens.

Appendix A

Camera intrinsic parameters estimation

The Imaged Circular Points (ICPs), which are the projections of the Circular Points are in the form of $\mathbf{e_\pm} = \mathbf{h_1} \pm i\mathbf{h_2}$ where $\mathbf{h_1}$ and $\mathbf{h_2}$ are the first two columns of the world-to-image homography $\mathbf{H}$ (Eq. 7). The Image of the Absolute Conic (IAC), denoted ω, encodes the intrinsic properties of the camera and is defined by

$$\omega \sim \mathbf{K}^{-T}\mathbf{K}^{-1} = \begin{bmatrix} \frac{1}{f_x^2} & 0 & \frac{-u_0}{f_x^2} \\ 0 & \frac{1}{f_y^2} & \frac{-v_0}{f_y^2} \\ \frac{-u_0}{f_x^2} & \frac{-v_0}{f_y^2} & \frac{u_0^2}{f_x^2} + \frac{v_0^2}{f_y^2} + 1 \end{bmatrix}$$

As stated in [15], all intrinsic calibration parameters can be obtained first by computing the image of the absolute conic (IAC) with precision from the imaged circular points using at least three images of two concentric circles under different orientations. By using this method, the intrinsic camera parameters in Eq. 5 can be estimated. If each ICP lies on the IAC, the following constraint is enforced:

$$\mathbf{e_\pm}^T \omega \mathbf{e_\mp} \sim 0$$

which can be rewritten as

$$\begin{aligned} \mathbf{h_1}^T \omega \mathbf{h_2} &= 0 \\ \mathbf{h_1}^T \omega \mathbf{h_1} - \mathbf{h_2}^T \omega \mathbf{h_2} &= 0 \end{aligned} \tag{26}$$

Each view gives two linear equations in the upper diagonal elements of the IAC [13]. If $\mathbf{K}$ is constant, three views are necessary to obtain a solution. As presented in [15], one ICP (and therefore its complex conjugate) can be recovered from one view of a pair of projected concentric circles $\{\mathbf{C_1}, \mathbf{C_2}\}$ computing the rank-2 matrix $\mathbf{\Delta_2}$:

$$\mathbf{\Delta_2} = \beta_2 \mathbf{C_1^{-1}} - \mathbf{C_2^{-1}} \sim \mathbf{H} \operatorname{diag}(0,\, 0,\, 1) \mathbf{H}^T$$

The ICPs are then recovered in the form of $\sqrt{s_1}\mathbf{u_1} \pm i\sqrt{s_2}\mathbf{u_2}$, where $\mathbf{u_1}$ and $\mathbf{u_2}$ are the first two columns of $\mathbf{U}$ resulting from the SVD (Single Value Decomposition) of $\Delta_2 = \mathbf{U}\operatorname{diag}(s_1, s_2, 0)\mathbf{U}^T$ with $\mathbf{U}\mathbf{U}^T = \mathbf{I}$. The determination of $\mathbf{K}$, is performed by estimating the IAC, ω, as the locus of all ICPs, through a linear formulation of the problem (Eq. 26). Finally, a Cholesky factorization of the IAC gives the matrix $\mathbf{K}$.

References

[1] Q. Zhang and R. Pless. Extrinsic calibration of a camera and laser range finder (improves camera calibration). In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2301–2306, Sendai, (2004).

[2] R. Dupont, R. Keriven, and P. Fuchs. An improved calibration technique for coupled single-row telemeter and ccd camera. In *Proc. The International Conference on 3-D Digital Imaging and Modeling*, pages 89–94, Ottawa, (2005).

[3] D. Scaramuzza, A. Harati, and R. Siegwart. Extrinsic self calibration of a camera and a 3d laser range finder from natural scenes. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4164–4169, San Diego, (2007).

[4] H. Zhao, L. Xiong, Z. Jiao, J. Cui, and H. Zha. Sensor alignment towards an omni-directional measurement using an intelligent vehicle. In *Proc. IEEE Intelligent Vehicles Symposium*, pages 292–298, Xi'an, (2009).

[5] C. Mei. *Laser-Augmented Omnidirectional Vision for 3D Localisation and Mapping*. PhD thesis, Ecole Nationale Superieure des Mines de Paris, (2007).

[6] L. Huang and M. Barth. A novel multi-planar lidar and computer vision calibration procedure using 2d patterns for automated navigation. In *Proc. IEEE Intelligent Vehicle Symposium*, pages 117–122, Xi'an, (2009).

[7] A. Willis, M. Zapata, and J. Conrad. A linear method for calibrating lidar-and-camera systems. In *Proc. IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS 09)*, pages 1–3, London, (2009).

[8] S. Bileschi. Fully automatic calibration of lidar and video streams from a vehicle. In *Proc. 3-D Digital Imaging and Modeling (3DIM ICCV Workshop)*, pages 1457–1464, Kyoto, (2009).

[9] O. Naroditsky, A. Patterson, and K. Daniilidis. Automatic alignment of a camera with a line scan lidar system. In *Proc. IEEE International Conference on Robotics and Automation*, pages 3429–3434, Shanghai, (2011).

[10] D. Klimentjew, N. Hendrich, and J. Zhang. Multi sensor fusion of camera and 3d laser range finder for object recognition. In *Proc. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 236–241, Salt Lake City, (2010).

[11] C. M. Shakarji. Least-squares fitting algorithms of the nist algorithm testing system. Technical Report 6, National Institute of Standards and Technology, (1998).

[12] D.W. Marquardt. An algorithm for the least-squares estimation of nonlinear parameters. *SIAM Journal of Applied Mathematics*, 11:431–441, (1963).

[13] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision. Second Edition.* Cambridge University Press, (2003).

[14] B. Prescott and G. McLean. Line-based correction of radial lens distortion. *Graphical Models and Image Processing*, 59:39–47, (1997).

[15] J.-S. Kim, P. Gurdjos, and I.-S. Kweon. Geometric and algebraic constraints of projected concentric circles and their applications to camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4), (2005).

[16] V. Fremont and R. Chellali. Direct camera calibration using two concentric circles from a single view. In *Proc. International Conference on Artificial Reality and Telexistence*, pages 93–98, Tokyo, (2002).

[17] K. Kanatani and N. Ohta. Automatic detection of circular objects by ellipse growing. *Int. J. Image Graphics*, 4(1):35–50, (2004).

[18] W. Gander, G. H. Golub, and R. Strebel. Least-squares fitting of circles and ellipses. Technical report, Eidgenossische Technische Hochschule, (1994).

[19] J.G. Semple and G. T. Kneebone. *Algebraic Projective Geometry.* Clarendon Press, Oxford Classic Series, (1952).

[20] K. Kanatani. *Geometric Computation for Machine Vision.* Oxford University Press, (1993).

[21] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):698–700, (1987).

[22] G. H. Golub and C. F. Van Loan. *Matrix computations (3rd ed.).* Johns Hopkins University Press, (1996).

[23] C. V. Stewart. Robust parameter estimation in computer vision. *Society for Industrial and Applied Mathematics*, 41(3):513–537, (1999).

[24] G. A. F. Seber and C. J. Wild. *Nonlinear Regression.* Wiley-Interscience Paperback Series, (2003).

[25] R. Y. Tsai. A versatile camera calibration techniaue for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4), (1987).

[26] S. A. Rodriguez, V. Fremont, P. Bonnifait, and V. Cherfaoui. Visual confirmation of mobile objects tracked by a multi-layer lidar. In *Proc. IEEE International Conference on Intelligent Transportation Systems*, Madeira, (2010).