Automatic Targetless Extrinsic Calibration of a 3D Lidar and Camera by Maximizing Mutual Information

Gaurav Pandey[1] and **James R. McBride**[2] and **Silvio Savarese**[1] and **Ryan M. Eustice**[3]

[1]Department of Electrical Engineering & Computer Science, University of Michigan, Ann Arbor, MI 48109, USA
[2]Research and Innovation Center, Ford Motor Company, Dearborn, MI 48124, USA
[3]Department of Naval Architecture & Marine Engineering, University of Michigan, Ann Arbor, MI 48109, USA

Abstract

This paper reports on a mutual information (MI) based algorithm for automatic extrinsic calibration of a 3D laser scanner and optical camera system. By using MI as the registration criterion, our method is able to work *in situ* without the need for any specific calibration targets, which makes it practical for in-field calibration. The calibration parameters are estimated by maximizing the mutual information obtained between the sensor-measured surface intensities. We calculate the Cramer-Rao-Lower-Bound (CRLB) and show that the sample variance of the estimated parameters empirically approaches the CRLB for a sufficient number of views. Furthermore, we compare the calibration results to independent ground-truth and observe that the mean error also empirically approaches to zero as the number of views are increased. This indicates that the proposed algorithm, in the limiting case, calculates a minimum variance unbiased (MVUB) estimate of the calibration parameters. Experimental results are presented for data collected by a vehicle mounted with a 3D laser scanner and an omnidirectional camera system.

1 Introduction

Today, robots are used to perform challenging tasks that we would not have imagined twenty years ago. In order to perform these complex tasks, robots need to sense and understand the environment around them. Depending upon the task at hand, robots are often equipped with different sensors to perceive their environment. Two important categories of perception sensors mounted on a robotic platform are: (i) range sensors (e.g., 3D/2D lidars, radars, sonars) and (ii) cameras (e.g., perspective, stereo, omnidirectional). Oftentimes the data obtained from these sensors is used independently; however, these modalities capture complementary information about the environment, which can be fused together by extrinsically calibrating the sensors. Extrinsic calibration is the process of estimating the rigid-body transformation between the reference (co-ordinate) system of the two sensors. This rigid-body transformation allows reprojection of the 3D points from the range sensor coordinate frame to the 2D camera coordinate frame (Fig. 1).

Substantial prior work has been done on extrinsic calibration of pinhole perspective cameras to 2D laser scanners
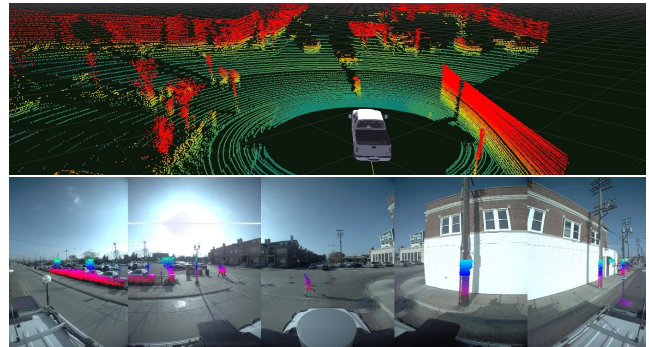
Figure 1: The top panel is a perspective view of the 3D lidar range data, color-coded by height above the ground plane. The bottom panel depicts the 3D lidar points projected onto the time-corresponding omnidirectional image. Several recognizable objects are present in the scene (people, stop signs, lamp posts, trees). (Only nearby objects are projected for visual clarity.)

(Zhang 2004; Mei and Rives 2006; Unnikrishnan and Hebert 2005) as they are inexpensive and are significantly helpful in many robotics applications. Zhang (2004) described a method that requires a planar checkerboard pattern to be simultaneously observed by the laser and camera systems. Mei and Rives (2006) later reported an algorithm for the calibration of a 2D laser range finder and an omnidirectional camera for both visible (i.e., laser is observed in camera image also) and invisible lasers.

2D laser scanners are used commonly for planar robotics applications, but recent advancements in 3D laser scanners have greatly extended the capabilities of robots. In most mobile robotics applications, the robot needs to automatically navigate and map the environment around them. In order to create realistic 3D maps, the 3D laser scanner and camera-system mounted on the robot need to be extrinsically calibrated. The problem of 3D laser to camera calibration was first addressed by Unnikrishnan and Hebert (2005), who extended Zhang's method (2004) to calibrate a 3D laser scanner with a perspective camera. Scaramuzza, Harati, and Siegwart (2007) later introduced a technique for the calibration of a 3D laser scanner and omnidirectional camera using

manual selection of point correspondences between camera and lidar. Aliakbarpour et al. (2009) proposed a technique for calibration of a 3D laser scanner and a stereo camera using an inertial measurement unit (IMU) to decrease the number of points needed for a robust calibration. Recently, Pandey et al. (2010) introduced a 3D lidar-camera calibration method that requires a planar checkerboard pattern to be viewed simultaneously from the laser scanner and camera system.

Here, we consider the automatic, targetless, extrinsic calibration of a 3D laser scanner and camera system. The attribute that no special targets need to be viewed makes the algorithm especially suitable for in-field calibration. To achieve this, the reported algorithm uses a mutual information (MI) framework based on the registration of the intensity and reflectivity information between the camera and laser modalities.

The idea of MI based multi-modal image registration was first introduced by Viola and Wells (1997) and Maes et al. (1997). Since then, the algorithmic developments in MI based registration have been exponential and have became state-of-the-art, especially in the medical image registration field. Within the robotics community, the application of MI has not been as widespread, even though robots today are often equipped with different modality sensors. Alempijevic et al. (2006) reported a MI based calibration framework that required a moving object to be observed in both sensor modalities. Because of their MI formulation, the results of Alempijevic et al. are (in a general sense) related to this work; however, their formulation of the MI cost-function ends up being entirely different due to their requirement of having to track moving objects. Boughorbal et al. (2000) proposed a χ^2 test that maximizes the statistical dependence of the data obtained from the two sensors for the calibration problem. This was later used by Williams et al. (2004) along with two methods to estimate an initial guess of the rigid-body transformation, which required manual intervention and a special object tracking mechanism. Boughorbal et al. (2000) and Williams et al. (2004) are the most closely related previous works to our own; however, they have reported problems of existence of local maxima in the cost-function formulated using either MI or χ^2 statistics.

In this work we solve this problem by incorporating scans from different scenes in a single optimization framework, thereby, obtaining a smooth and concave cost function, easy to solve by any gradient ascent algorithm. Fundamentally, we can use either MI or the χ^2 test as both of them provide a measure of statistical dependence of the two random variables (McDonald 2009). We chose MI because of ongoing active research in fast and robust MI estimation techniques, such as James-Stein-type shrinkage estimators (Hausser and Strimmer 2009), which have the potential to be directly employed in the proposed framework, though are not currently. Importantly, in this work we provide a measure of the uncertainty of the estimated calibration parameters and empirically show that it achieves the Cramer-Rao-Lower-Bound, indicating that it is an efficient estimator.
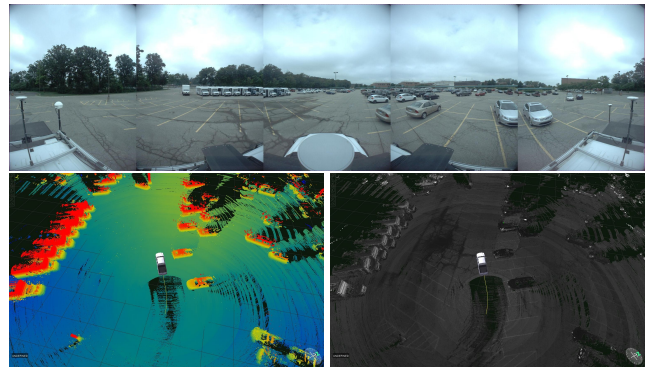

Figure 2: The top panel is an image from the Ladybug3 omnidirectional camera. The bottom panel depicts the Velodyne-64E 3D lidar data color-coded by height (left), and by laser reflectivity (right).
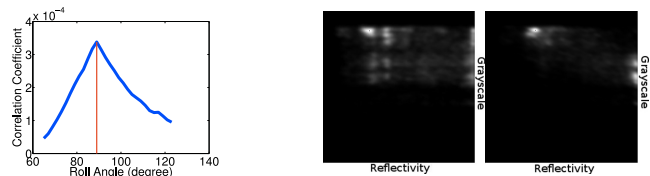

Figure 3: The left panel shows the correlation coefficient as a function of one of the rotation parameters (keeping all other parameters fixed at their true value). We observe that the correlation coefficient is maximum for the true roll angle of 89°. Depicted in the right panel is the joint histogram of the reflectivity and the intensity values when calculated at an incorrect (left) and correct (right) transformation. Note that the joint histogram is least dispersed under the correct transformation.

2   Methodology

In our work we have used a Velodyne 3D laser scanner (Velodyne 2007) and a Ladybug3 omnidirectional camera system (Pointgrey 2009) mounted to the roof of a vehicle. A snapshot of the type of data that we obtain from these sensors is depicted in Fig. 2, and clearly exhibits visual correlation between the two modalities. We assume that the intrinsic calibration parameters of both the camera system and laser scanner are known. We also assume that the laser scanner reports meaningful surface reflectivity values. In this work, we have previously calibrated the reflectivity values of the laser scanner using the algorithm reported by Levinson and Thrun (2010).

Our claim about the correlation between the laser reflectivity and camera intensity values is verified by a simple experiment shown in Fig. 3. Here we calculate the correlation coefficient for the reflectivity and intensity values for Fig. 2's scan-image pair at different values of the calibration parameter and observe a distinct maxima at the true value. Moreover, in the right panel we observe that the joint histogram of the laser reflectivity and the camera intensity values is least dispersed when calculated under the correct transformation parameters.

Figure 4: (left) Image with shadows of trees and buildings on the road. (right) Top view of the corresponding lidar reflectivity map, which is unaffected by ambient lighting.

Although scenarios such as Fig. 2 do exhibit high correlation between the two sensors, there exist other scenarios where they might not be as strongly correlated. One such example is shown in Fig. 4. Here, the ambient light plays a critical role in determining the intensity levels of the image pixels. As clearly depicted in the image, there are some regions of the road that are covered by shadows. The gray levels of the image are affected by the shadows; however, the corresponding reflectivity values in the laser are not because it uses an active lighting principle. Thus, in these type of scenarios the data between the two sensors might not show as strong of a correlation and, hence, will produce a weak input for the proposed algorithm. In this paper, we do not focus on solving the general lighting problem. Instead, we formulate a MI based data fusion criterion to estimate the extrinsic calibration parameters between the two sensors assuming that the data is, for the most part, not corrupted by lighting artifacts. In fact, for many practical indoor/outdoor calibration scenes (e.g., Fig. 2) shadow effects represent a small fraction of the overall data and thus appear as noise in the calibration process. This is easily handled by the proposed method by aggregating multiple views.

2.1 Theory

The mutual information (MI) between two random variables X and Y is a measure of the statistical dependence occurring between the two random variables. Various formulations of MI have been presented in the literature, each of which demonstrate a measure of statistical dependence of the random variables in consideration. One such form of MI is defined in terms of entropy of the random variables:

$$\mathrm{MI}(X, Y) = \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(X, Y), \quad (1)$$

where $\mathrm{H}(X)$ and $\mathrm{H}(Y)$ are the entropies of random variables X and Y, respectively, and $\mathrm{H}(X, Y)$ is the joint entropy of the two random variables:

$$\mathrm{H}(X) = -\sum_{x \in X} p_X(x) \log p_X(x), \quad (2)$$

$$\mathrm{H}(Y) = -\sum_{y \in Y} p_Y(y) \log p_Y(y), \quad (3)$$

$$\mathrm{H}(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p_{XY}(x, y) \log p_{XY}(x, y). \quad (4)$$

The entropy $\mathrm{H}(X)$ of a random variable X denotes the amount of uncertainty in X, whereas $\mathrm{H}(X, Y)$ is the amount of uncertainty when the random variables X and Y are co-observed. Hence, (1) shows that $\mathrm{MI}(X, Y)$ is the reduction in the amount of uncertainty of the random variable X when we have some knowledge about random variable Y. In other words, $\mathrm{MI}(X, Y)$ is the amount of information that Y contains about X and vice versa.

2.2 Mathematical Formulation

Here we consider the laser reflectivity value of a 3D point and the corresponding grayscale value of the image pixel to which this 3D point is projected as the random variables X and Y, respectively. The marginal and joint probabilities of these random variable $p(X)$, $p(Y)$ and $p(X, Y)$ can be obtained from the normalized marginal and joint histograms of the reflectivity and grayscale intensity values of the 3D points co-observed by the laser scanner and camera. Let $\{\mathbf{P}_i; i = 1, 2, \cdots, n\}$ be the set of 3D points whose coordinates are known in the laser reference system and let $\{X_i; i = 1, 2, \cdots, n\}$ be the corresponding reflectivity values for these points ($X_i \in [0, 255]$).

For the usual pinhole camera model, the relationship between a homogeneous 3D point, $\tilde{\mathbf{P}}_i$, and its homogeneous image projection, $\tilde{\mathbf{p}}_i$, is given by:

$$\tilde{\mathbf{p}}_i = \mathrm{K}[\mathrm{R} \,|\, \mathbf{t}]\tilde{\mathbf{P}}_i, \quad (5)$$

where $(\mathrm{R}, \mathbf{t})$, called the extrinsic parameters, are the orthonormal rotation matrix and translation vector that relate the laser coordinate system to the camera coordinate system, and K is the camera intrinsic matrix. Here R is parametrized by the Euler angles $[\phi, \theta, \psi]^\top$ and $\mathbf{t} = [x, y, z]^\top$ is the Euclidean 3-vector. Let $\{Y_i; i = 1, 2, \cdots, n\}$ be the grayscale intensity value of the image pixel upon which the 3D point projects such that

$$Y_i = I(\mathbf{p}_i), \quad (6)$$

where $Y_i \in [0, 255]$ and I is the grayscale image.

Thus, for a given set of extrinsic calibration parameters, X_i and Y_i are the observations of the random variables X and Y, respectively. The marginal and joint probabilities of the random variables X and Y can be obtained from the kernel density estimate (KDE) of the normalized marginal and joint histograms of X_i and Y_i. The KDE of the joint distribution of the random variables X and Y is given by (Scott 1992):

$$p(X, Y) = \frac{1}{n} \sum_{i=1}^{n} K_\Omega \left(\begin{bmatrix} X \\ Y \end{bmatrix} - \begin{bmatrix} X_i \\ Y_i \end{bmatrix} \right), \quad (7)$$

where $K(\cdot)$ is the symmetric kernel and Ω is the *bandwidth* or the *smoothing* matrix of the kernel. In our experiments we have used a Gaussian kernel and a bandwidth matrix Ω proportional to the square root of the sample covariance matrix ($\Sigma^{1/2}$) of the data. An illustration of the KDE of the probability distribution of the grayscale values from the available histograms is shown in Fig. 5.

Once we have an estimate of the probability distribution we can write the MI of the two random variables as a function of the extrinsic calibration parameters $(\mathrm{R}, \mathbf{t})$, thereby formulating an objective function:

$$\hat{\mathbf{\Theta}} = \arg \max_{\mathbf{\Theta}} \ \mathrm{MI}(X, Y; \mathbf{\Theta}), \quad (8)$$
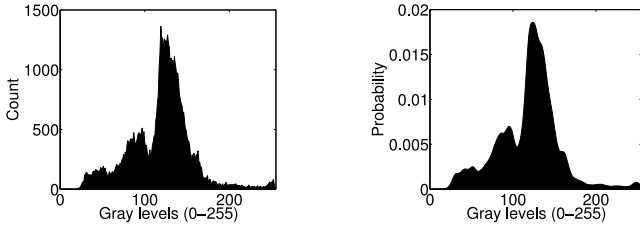
Figure 5: Kernel density estimate of the probability distribution (right), estimated from the observed histogram (left) of grayscale intensity values.

whose maxima occurs at the sought after calibration parameters, $\mathbf{\Theta} = [x, y, z, \phi, \theta, \psi]^\top$.

2.3 Optimization

We use the Barzilai-Borwein (BB) steepest gradient ascent algorithm (Barzilai and Borwein 1988) to find the calibration parameters $\mathbf{\Theta}$ that maximizes (8). The BB method proposes an adaptive step size in the direction of the gradient of the cost function. The step size incorporates the second order information of the objective function. If the gradient of the cost function (8) is given by:

$$\mathbf{G} \equiv \nabla\,\mathrm{MI}(X, Y; \mathbf{\Theta}), \qquad (9)$$

then one iteration of the BB method is defined as:

$$\mathbf{\Theta}_{k+1} = \mathbf{\Theta}_k + \gamma_k \frac{\mathbf{G}_k}{\|\mathbf{G}_k\|}, \qquad (10)$$

where $\mathbf{\Theta}_k$ is the optimal solution of (8) at the k-th iteration, $\mathbf{G}_k$ is the gradient vector (computed numerically) at $\mathbf{\Theta}_k$, $\|\cdot\|$ is the Euclidean norm and γ_k is the adaptive step size, which is given by:

$$\gamma_k = \frac{\mathbf{s}_k^\top \mathbf{s}_k}{\mathbf{s}_k^\top \mathbf{g}_k}, \qquad (11)$$

where $\mathbf{s}_k = \mathbf{\Theta}_k - \mathbf{\Theta}_{k-1}$ and $\mathbf{g}_k = \mathbf{G}_k - \mathbf{G}_{k-1}$.

The convex nature of the cost function (Fig. 6) is achieved by aggregating scans from different scenes in a single optimization framework and allows the algorithm to converge to the global maximum in a few steps. Typically the algorithm takes around 2-10 minutes to converge, depending upon the number of scans used to estimate MI. The complete algorithm is shown in Algorithm 1.

2.4 Cramer-Rao-Lower-Bound of the Variance of the Estimated Parameters

It is important to know the uncertainty in the estimated parameters in order to use them in any vision or simultaneous localization and mapping (SLAM) algorithm. Here we use the Cramer-Rao-Lower-Bound (CRLB) of the variance of the estimated parameters as a measure of the uncertainty. The CRLB (Cramer 1946) states that the variance of any unbiased estimator is greater than or equal to the inverse of the Fisher Information matrix. Moreover, any unbiased estimator that achieves this lower bound is said to be efficient. The Fisher information of a random variable Z is a measure of
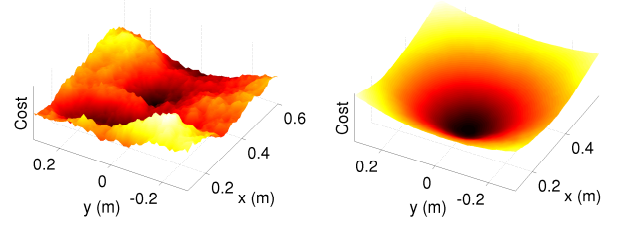

Figure 6: The MI cost-function surface versus translation parameters x and y for a single scan (left) and aggregation of 10 scans (right). Note the global convexity and smoothness when the scans are aggregated. The correct value of parameters is given by (0.3, 0.0). Negative MI is plotted here to make visualization of the extrema easier.

Algorithm 1 Automatic Calibration by maximization of MI

1: **Input:** 3D Point cloud $\{\mathbf{P}_i; i = 1, 2, \cdots, n\}$, Reflectivity $\{X_i; i = 1, 2, \cdots, n\}$, Image $\{I\}$, Initial guess $\{\mathbf{\Theta}_0\}$.
2: **Output:** Estimated parameter $\{\hat{\mathbf{\Theta}}\}$.
3: **while** $(\|\mathbf{\Theta}_{k+1} - \mathbf{\Theta}_k\| > THRESHOLD)$ **do**
4:    $\mathbf{\Theta}_k \rightarrow [\mathrm{R}\,|\,\mathbf{t}]$
5:    **for** $i = 1 \rightarrow n$ **do**
6:      $\tilde{\mathbf{p}}_i = \mathrm{K}[\mathrm{R}\,|\,\mathbf{t}]\tilde{\mathbf{P}}_i$
7:      $Y_i = I(\mathbf{p}_i)$
8:    **end for**
9:    Calculate the joint histogram: $\mathrm{Hist}(X, Y)$.
10:    Calculate the kernel density estimate of the joint distribution: $p(X, Y; \mathbf{\Theta}_k)$.
11:    Calculate the MI: $\mathrm{MI}(X, Y; \mathbf{\Theta}_k)$.
12:    Calculate the gradient: $\mathbf{G}_k = \nabla\,\mathrm{MI}(X, Y; \mathbf{\Theta}_k)$.
13:    Calculate the step size γ_k.
14:    $\mathbf{\Theta}_{k+1} = \mathbf{\Theta}_k + \gamma_k \frac{\mathbf{G}_k}{\|\mathbf{G}_k\|}$.
15: **end while**

---

the amount of information that the observations of the random variable Z carries about an unknown parameter α, on which the probability distribution of Z depends. If the distribution of a random variable Z is given by $f(Z; \alpha)$ then the Fisher information is given by (Lehmann and Casella 2011):

$$\mathcal{I}(\alpha) = \mathrm{E}\left[\left(\frac{\partial}{\partial\alpha}\log f(Z; \alpha)\right)^2\right]. \qquad (12)$$

In our case the joint distribution of the random variables X and Y (as defined in (7)) depends upon the six dimensional transformation parameter $\mathbf{\Theta}$. Therefore, the Fisher information is given by a $[6 \times 6]$ matrix

$$\mathcal{I}(\mathbf{\Theta})_{ij} = \mathrm{E}\left[\frac{\partial}{\partial\mathbf{\Theta}_i}\log p(X, Y; \mathbf{\Theta})\frac{\partial}{\partial\mathbf{\Theta}_j}\log p(X, Y; \mathbf{\Theta})\right], \qquad (13)$$

and the required CRLB is given by

$$\mathrm{Cov}(\mathbf{\Theta}) \leq \mathcal{I}(\mathbf{\Theta})^{-1}, \qquad (14)$$

where $\mathcal{I}(\mathbf{\Theta})^{-1}$ is the inverse of the Fisher information matrix calculated at the estimated value of the parameter $\hat{\mathbf{\Theta}}$.

3 Experiments and Results

We present results from real data collected from a 3D laser scanner (a Velodyne HDL-64E) and an omnidirectional camera system (a Point Grey Ladybug3) mounted on the roof of a vehicle. Although we present results from an omnidirectional camera system, the algorithm is applicable to any kind of laser-camera system, including monocular imagery. In all of our experiments *scan* refers to a single 360° field of view 3D point cloud and its time-corresponding camera imagery.

3.1 Calibration Performance Using a Single Scan

In this experiment we show that the *in situ* calibration performance is dependent upon the environment in which the scans are collected. We collected several datasets in both indoor and outdoor settings. The indoor dataset was collected inside a large garage, and exhibited many nearby objects such as walls and other vehicles. In contrast, most of the outdoor dataset did not have many close by objects. In the absence of near-field 3D points, the cost-function is insensitive to the translational parameters—making them more difficult to estimate. This is a well-known phenomenon of projective geometry, where in the limiting case if we consider points at infinity, $[\tilde{x}, \tilde{y}, \tilde{z}, 0]^\top$, the projection of these points (also known as the vanishing points) are not affected by the translational component of the camera projection matrix. Hence, we should expect that scans that only contain 3D points far-off in the distance (i.e., the outdoor dataset) will have poor observability of the extrinsic translation vector, $\mathbf{t}$, as opposed to scans that contain many nearby 3D points (i.e., the indoor dataset). In Fig. 7(a) and (b) we have plotted the calibration results for 15 scans collected in outdoor and indoor settings, respectively. We clearly see that the variability in the estimated parameters for the outdoor scans is much larger than that of the indoor scans. Thus, from this experiment we conclude that we need to have nearby objects in order to robustly estimate the calibration parameters from a single-view.

3.2 Calibration Performance Using Multiple Scans

In the previous section we showed that it is necessary to have nearby objects in the scans in order to robustly estimate the calibration parameters; however, this might not always be practical—depending on the environment. In this experiment we demonstrate improved calibration convergence by simply aggregating multiple scans into a single batch optimization process. Fig. 7(c) shows the calibration results for when multiple scans are considered in the MI calculation. In particular, the experiments show that the standard deviation of the estimated parameters quickly decreases as the number of scans are increased by just a few. Here, the red plot shows the standard deviation (σ) of the calibration parameters computed over 1000 trials, where in each trial we randomly sampled $\{N = 5, 10, \cdots, 40\}$ scans from the available indoor and outdoor datasets to use in the MI calculation. The green plot shows the corresponding CRLB of the standard deviation of the estimated parameters. In particular, we

see that with as little as 10–15 scans, we can achieve very accurate performance. Moreover, we see that the sample variance asymptotically approaches the CRLB as the number of scans are increased, indicating this is an efficient estimator.

3.3 Quantitative Verification of the Calibration Result

We performed the following three experiments to quantitatively verify the results obtained from the proposed method.

Comparison with χ^2 test (Williams et al. 2004) In this experiment we replace the MI criteria by the χ^2 statistic used by Williams et al. (2004). The χ^2 statistic gives a measure of the statistical dependence of the two random variables in terms of the closeness of the observed joint distribution to the distribution obtained by assuming X and Y to be statistically independent:

$$\chi^2(X, Y; \boldsymbol{\Theta}) = \sum_{x \in X, y \in Y} \frac{\left(p(x, y; \boldsymbol{\Theta}) - p(x; \boldsymbol{\Theta}) p(y; \boldsymbol{\Theta}) \right)^2}{p(x; \boldsymbol{\Theta}) p(y; \boldsymbol{\Theta})}.$$
(15)

We can therefore modify the cost function given in (8) to:

$$\boldsymbol{\Theta} = \arg\max_{\boldsymbol{\Theta}} \chi^2(X, Y; \boldsymbol{\Theta}).$$
(16)

The comparison of the calibration results obtained from the χ^2 test (16) and with the MI (8) (using 40 scan-image pairs) is shown in Table 1. We see that the results obtained from the χ^2 statistics are similar to those obtained from the MI criteria. This is mainly because the χ^2 statistics and MI are equivalent and essentially capture the amount of correlation between the two random variables (McDonald 2009). Moreover, aggregating several scans in a single optimization framework generates a smooth cost function, allowing us to completely avoid the estimation of the initial guess of the calibration parameters by manual methods introduced in (Williams et al. 2004).

Comparison with the checkerboard pattern method (Pandey et al. 2010) Pandey et al. proposed a method that requires a planar checkerboard pattern to be observed simultaneously from the laser scanner and the camera system. We compared our minimum variance results (i.e., estimated using 40 scans) with the results obtained from the method described in (Pandey et al. 2010) and found that they are very close (Table 1). The reprojection of 3D points on the image using results obtained from these methods look very similar visually. Therefore, in the absence of ground truth, it is difficult to say which result is more accurate. The proposed method though, is definitely much faster and easier as it does not involve any manual intervention.

Comparison with ground-truth from the omnidirectional camera's intrinsics The omnidirectional camera used in our experiments is pre-calibrated from the manufacturer. It has six 2-Megapixel cameras, with five cameras positioned in a horizontal ring and one positioned vertically, such that the rigid-body transformation of each camera with respect to a common coordinate frame, called the camera head, is well known. Here, $\mathbf{X}_{hc_i}$ is the Smith, Self,

(a) Single-scan: outdoor dataset.   (b) Single-scan: indoor dataset.   (c) Multi-scan: outdoor and indoor dataset.
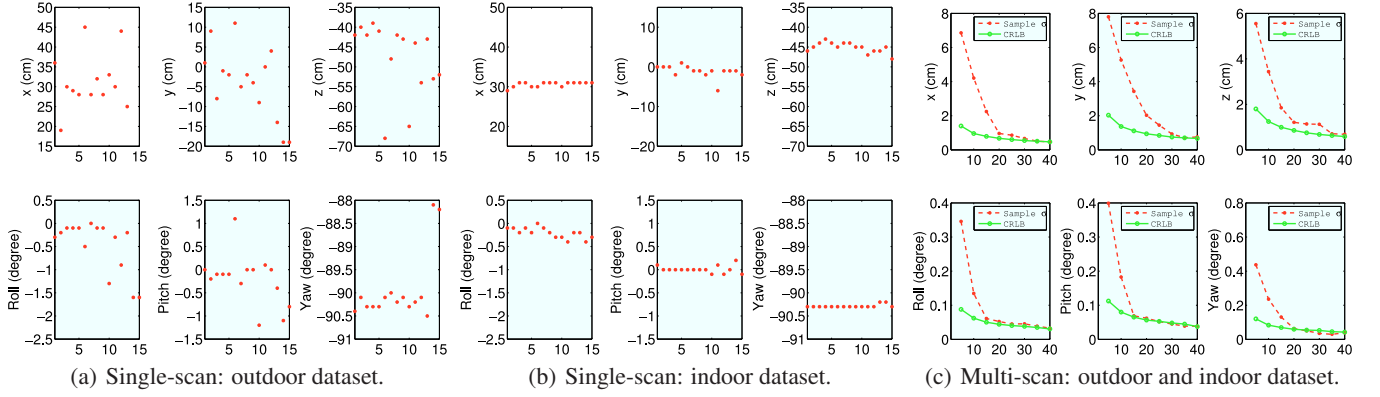
Figure 7: Single-view calibration results for outdoor and indoor datasets are shown in (a), (b). The variability in the estimated parameters (especially translation) is significantly larger in the case of the outdoor dataset. Each point on the abscissa corresponds to a different trial (i.e., different scan). Multiple-view calibration results are shown in (c). Here we use all five (horizontal) images from the Ladybug3 omnidirectional camera during the calibration. Plotted is the uncertainty of the recovered calibration parameter versus the number of scans used. The red plot shows the sample-based standard deviation (σ) of the estimated calibration parameters calculated over 1000 trials. The green plot represents the corresponding CRLB of the standard deviation of the estimated parameters. Each point on the abscissa corresponds to the number of aggregated scans used per trial.

Table 1: Comparison of calibration parameters estimated by: [a] (proposed method), [b] (Williams et al. 2004), [c] (Pandey et al. 2010).

	x [cm]	y [cm]	z [cm]	Roll [deg]	Pitch [deg]	Yaw [deg]
a	30.5	-0.5	-42.6	-0.15	0.00	-90.27
b	29.8	0.0	-43.4	-0.15	0.00	-90.32
c	34.0	1.0	-41.6	0.01	-0.03	-90.25

and Cheeseman (1988) coordinate frame notation, and represents the 6-DOF pose ($\mathbf{X}_{hc_i}$) of the ith camera (c_i) with respect to the camera head (h). Since we know $\mathbf{X}_{hc_i}$ from the manufacturer, we can calculate the pose of the ith camera with respect to the jth camera as:

$$\mathbf{X}_{c_i c_j} = \ominus \mathbf{X}_{hc_i} \oplus \mathbf{X}_{hc_j}, \quad \{i \neq j\}. \quad (17)$$

In the previous experiments we used all 5 horizontally positioned cameras of the Ladybug3 omnidirectional camera system to calculate the MI; however, in this experiment we consider only one camera at a time and directly estimate the pose of the camera with respect to the laser reference frame ($\mathbf{X}_{\ell c_i}$). This allows us to calculate $\hat{\mathbf{X}}_{c_i c_j}$ from the estimated calibration parameters $\hat{\mathbf{X}}_{\ell c_i}$. Thus, we can compare the true value of $\mathbf{X}_{c_i c_j}$ (from the manufacturer data) with the estimated value $\hat{\mathbf{X}}_{c_i c_j}$.

Fig. 8 shows one such comparison from the two side looking cameras of the Ladybug3 camera system. Here we see that the error in the estimated calibration parameters reduces with the increase in the number of scans and asymptotically approaches the expected value of the error (i.e., $\mathrm{E}[|\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}|] \to 0$). It should be noted that in this experiment we used only a single camera as opposed to all 5 cameras of the omnidirectional camera system, thereby reducing the
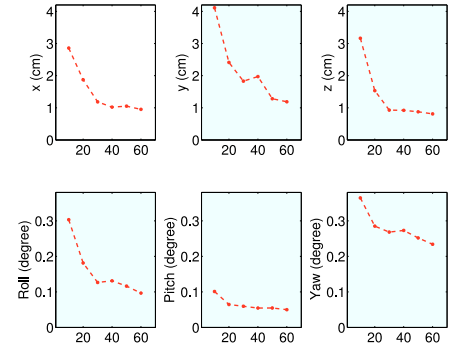

Figure 8: Comparison with ground-truth. Here we have plotted the mean absolute error in the calibration parameters ($|\mathbf{X}_{c_i c_j} - \hat{\mathbf{X}}_{c_i c_j}|$) versus the number of scans used to estimate these parameters. The mean is calculated over 100 trials of sampling N, where $N = 10, 20, \cdots, 60$ scans per trial. We see that the error decreases as the number of scans are increased.

amount of data used in each trial to $1/5^{th}$. It is our conjecture that with additional trials, a statistically significant validation of unbiasedness could be achieved. Since the sample variance of the estimated parameters also approaches the CRLB as the number of scans are increased, in the limit our estimator should exhibit the properties of a MVUB estimator (i.e., in the limiting case the CRLB can be considered as the true variance of the estimated parameters). Since in this experiment we have used only one camera of the omnidirectional camera system to estimate the calibration parameter, we have demonstrated that the proposed method can be used for any standard laser-camera system (i.e., monocular too).

4 Conclusions and Future works

This paper reported an information theoretic algorithm to automatically estimate the rigid-body transformation between a camera and 3D laser scanner by exploiting the statistical dependence between the two measured modalities. In this work MI was chosen as the measure of this statistical dependence. The most important thing to take away about this algorithm is that it is completely data driven and does not require any artificial targets to be placed in the field-of-view of the sensors.

Generally, sensor calibration in a robotic application is performed once, and the same calibration is assumed to be true for rest of the life of that particular sensor suite. However, for robotics applications where the robot needs to go out into rough terrain, assuming that the sensor calibration is not altered during a task is often not true. Although, we should calibrate the sensors before every task, it is typically not practical to do so if it requires to setup a calibration environment every time. Our method, being free from any such constraints, can be easily used to fine tune the calibration of the sensors *in situ*, which makes it applicable to in-field calibration scenarios. Moreover, our algorithm provides a measure of the uncertainty of the estimated parameters through the CRLB.

Future works will explore the incorporation of other sensing modalities (e.g., sonars or laser without reflectivity) into the proposed framework. We believe that even if the sensor modalities do not provide a direct correlation (as observed between reflectivity and grayscale values), one can extract similar features from the two modalities, which can be used in the MI framework. For instance, if the lidar just gives the range returns (no reflectivity), then we can first generate a depth map from the point cloud. The depth map and the corresponding image should both have edge and corner features at the discontinuities in the environment. The MI between these features should exhibit a maxima at the sought after rigid-body transformation.

Acknowledgments

References

Alempijevic, A.; Kodagoda, S.; Underwood, J. P.; Kumar, S.; and Dissanayake, G. 2006. Mutual information based sensor registration and calibration. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, 25–30.

Aliakbarpour, H.; Nunez, P.; Prado, J.; Khoshhal, K.; and Dias, J. 2009. An efficient algorithm for extrinsic calibration between a 3d laser range finder and a stereo camera for surveillance. In *Int. Conf. on Advanced Robot.*, 1–6.

Barzilai, J., and Borwein, J. M. 1988. Two-point step size gradient methods. *IMA J. Numerical Analysis* 8:141–148.

Boughorbal, F.; Page, D. L.; Dumont, C.; and Abidi, M. A. 2000. Registration and integration of multisensor data for photorealistic scene reconstruction. In *Proc. of SPIE*, volume 3905, 74–84.

Cramer, H. 1946. *Mathematical methods of statistics*. Princeton landmarks in mathematics and physics. Princeton University Press.

Hausser, J., and Strimmer, K. 2009. Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. *J. Mach. Learning Res.* 10:1469–1484.

Lehmann, E. L., and Casella, G. 2011. *Theory of Point Estimation*. Springer Texts in Statistics Series. Springer.

Levinson, J., and Thrun, S. 2010. Unsupervised calibration for multi-beam lasers. In *Proc. Int. Symp. Experimental Robot.*

Maes, F.; Collignon, A.; Vandermeulen, D.; Marchal, G.; and Suetens, P. 1997. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imag.* 16:187–198.

McDonald, J. H. 2009. *Handbook of Biological Statistics*. Baltimore, MD USA: Sparky House Publishing, 2nd edition.

Mei, C., and Rives, P. 2006. Calibration between a central catadioptric camera and a laser range finder for robotic applications. In *Proc. IEEE Int. Conf. Robot. and Automation*, 532–537.

Pandey, G.; McBride, J. R.; Savarese, S.; and Eustice, R. M. 2010. Extrinsic calibration of a 3d laser scanner and an omnidirectional camera. In *IFAC Symp. Intell. Autonomous Vehicles*, volume 7.

Pointgrey. 2009. Spherical vision products: Ladybug3. Specification sheet and documentations available at www.ptgrey.com/products/ladybug3/index.asp.

Scaramuzza, D.; Harati, A.; and Siegwart, R. 2007. Extrinsic self calibration of a camera and a 3d laser range finder from natural scenes. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, 4164–4169.

Scott, D. W. 1992. *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York: John Wiley.

Smith, R.; Self, M.; and Cheeseman, P. 1988. A stochastic map for uncertain spatial relationships. In *Proc. Int. Symp. Robot. Res.*, 467–474. Santa Clara, CA USA: MIT Press.

Unnikrishnan, R., and Hebert, M. 2005. Fast extrinsic calibration of a laser rangefinder to a camera. Technical Report CMU-RI-TR-05-09, Robotics Institute Carnegie Mellon University.

Velodyne. 2007. Velodyne HDL-64E: A high definition LIDAR sensor for 3D applications. Available at www.velodyne.com/lidar/products/white_paper.

Viola, P., and Wells, W. 1997. Alignment by maximization of mutual information. *Int. J. Comput. Vis.* 24:137–154.

Williams, N.; Low, K. L.; Hantak, C.; Pollefeys, M.; and Lastra, A. 2004. Automatic image alignment for 3d environment modeling. In *Proc. IEEE Brazilian Symp. Comput. Graphics and Image Process.*, 388–395.

Zhang, Q. 2004. Extrinsic calibration of a camera and laser range finder. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, 2301–2306.