Automatic Extrinsic Calibration of Vision and Lidar by Maximizing Mutual Information

Gaurav Pandey
Electrical Engineering: Systems, University of Michigan, Ann Arbor, Michigan 48109
e-mail: pgaurav@umich.edu
James R. McBride
Research and Innovation Center, Ford Motor Company, Dearborn, Michigan 48124
e-mail: jmcbride@ford.com
Silvio Savarese
Computer Science Department, Stanford University, Stanford, California 94305
e-mail: ssilvio@stanford.edu
Ryan M. Eustice
Naval Architecture and Marine Engineering, University of Michigan, Ann Arbor, Michigan 48109
e-mail: eustice@umich.edu

Received 12 August 2013; accepted 25 June 2014

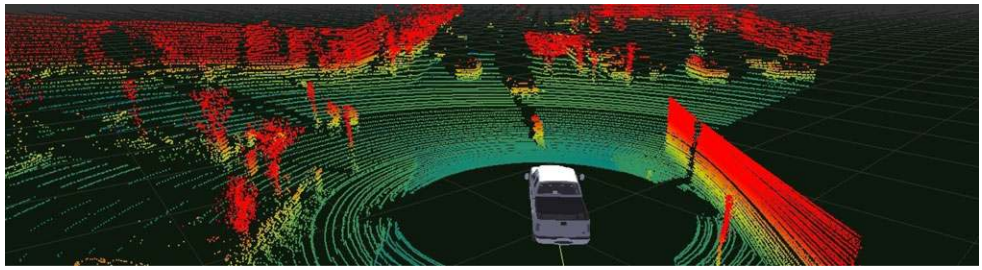
This paper reports on an algorithm for automatic, targetless, extrinsic calibration of a lidar and optical camera system based upon the maximization of mutual information between the sensor-measured surface intensities. The proposed method is completely data-driven and does not require any fiducial calibration targets—making *in situ* calibration easy. We calculate the Cramér-Rao lower bound (CRLB) of the estimated calibration parameter variance, and we show experimentally that the sample variance of the estimated parameters empirically approaches the CRLB when the amount of data used for calibration is sufficiently large. Furthermore, we compare the calibration results to independent ground-truth (where available) and observe that the mean error empirically approaches zero as the amount of data used for calibration is increased, thereby suggesting that the proposed estimator is a minimum variance unbiased estimate of the calibration parameters. Experimental results are presented for three different lidar-camera systems: (i) a three-dimensional (3D) lidar and omnidirectional camera, (ii) a 3D time-of-flight sensor and monocular camera, and (iii) a 2D lidar and monocular camera.

1. INTRODUCTION

With recent advancements in sensing technologies, the ability to equip a robot with multi-sensor lidar/camera configurations has greatly improved. Two important categories of perception sensors commonly mounted on a robotic platform are (i) range sensors [e.g., three-dimensional/two-dimensional (3D/2D) lidars, radars, sonars] and (ii) optical cameras (e.g., perspective, stereo, omnidirectional). Oftentimes the data obtained from these sensors are used independently; however, these modalities capture complementary information about the environment, which can be co-registered by extrinsically calibrating the sensors.

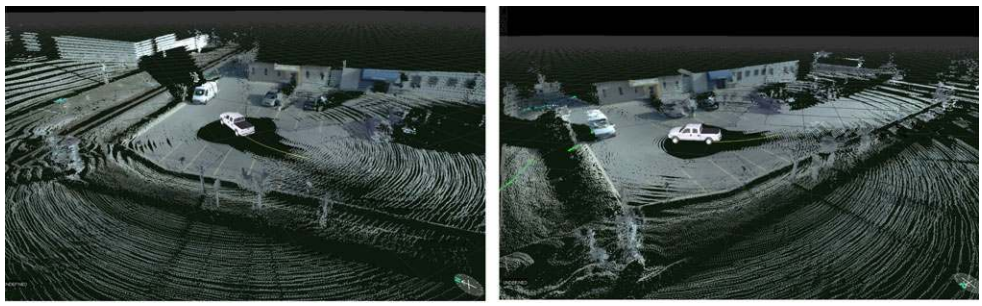
Extrinsic calibration is the process of estimating the rigid-body transformation between the reference coordinate system of the two sensors. This rigid-body transformation allows reprojection of the 3D points from the range sensor coordinate frame to the 2D camera coordinate frame (Figure 1). Fusion of data provided by range and vision sensors can enhance various state-of-the-art computer vision and robotics algorithms. For example, Bao and Savarese (2011) have proposed a novel framework for structure-from-motion (SFM) that takes advantage of both semantic (from camera data) and geometrical properties (from lidar data) associated with the objects in the scene. Pandey et al. (2011a) use the coregistered 3D point cloud with the camera imagery to bootstrap the scan registration process. They show that the incorporation of image data in the 3D scan registration process allows for robust registration without any initial guess (e.g., from odometry). Additionally, Pandey et al. (2012b) also proposed a robust mutual-information (MI) -based framework for incorporating coregistered camera and lidar data into the scan registration process. In mobile robotics, simultaneous localization and mapping (SLAM) is one of the basic tasks performed by robots. Although using a lidar for pose estimation and a camera for loop closure detection is common practice in SLAM (Newman

Direct correspondence to: Gaurav Pandey, e-mail: pgaurav@umich.edu



(a) 3D lidar point cloud

(b) Omnidirectional image with a subset of lidar points projected



(c) Fused RGB textured point cloud

Figure 1. Reprojection of lidar and camera via extrinsic rigid-body calibration. (a) Perspective view of the 3D lidar range data, color-coded by height above the ground plane. (b) Depiction of the 3D lidar points projected onto the time-corresponding omnidirectional camera image. Several recognizable objects are present in the scene (e.g., people, stop signs, lamp posts, trees). Only nearby objects are projected for visual clarity. (c) Depiction of two different views of a fused lidar/camera textured point cloud. Each 3D point is colored by the RGB value of the pixel corresponding to the projection of the point onto the image.

et al., 2006), several successful attempts have been made to use the coregistered data in the SLAM framework directly. Carlevaris-Bianco et al. (2011) proposed a novel mapping and localization framework that uses the co-registered omnidirectional camera imagery and lidar data to construct a map containing only the most viewpoint-robust visual features and then uses a monocular camera alone for online localization within the *a priori* map. Tamjidi and Ye (2012) reported a six degree of freedom (DOF) vehicle pose estimation algorithm that uses the fusion of lidar and camera data in both the feature initialization and motion prediction stages of an extended Kalman filter (EKF).

Extrinsic calibration is a core pre-requisite for gathering useful data from a multi-sensor platform. Many of the existing algorithms for extrinsic calibration of lidar-camera systems require that fiducial targets be placed in the field of view of the two sensors. A planar checkerboard pattern (Figure 2) is the most common calibration target used by researchers, as it is easy to extract from both camera and lidar data. The correspondences between lidar and camera data (e.g., point-to-point or point-to-plane) are established either manually or automatically, and calibration parameters are estimated by minimizing a reprojection error. The accuracy of these methods is dependent upon the accuracy of the established correspondences. There are also methods that do not require any special targets (Moghadam et al., 2013; Scaramuzza et al., 2007), but rely upon extraction of some features (e.g., edges, lines, corners) from the camera

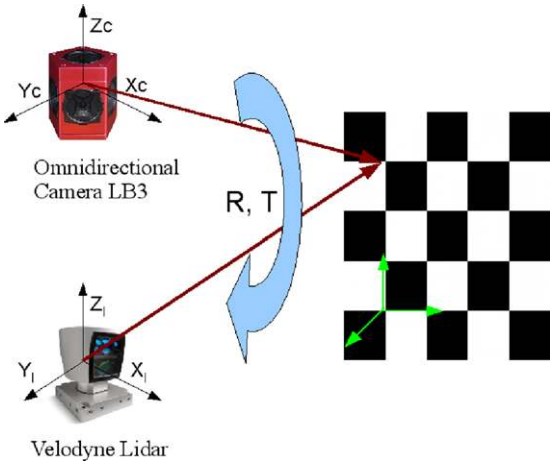


Figure 2. Typical target-based calibration setup for an omnidirectional camera and a 3D lidar using a planar checkerboard pattern.

and lidar data, either manually or automatically. The automatic feature extraction methods are generally not robust and require manual supervision to achieve small calibration errors. Although these methods can provide a good estimate of the calibration parameters, they are generally laborious and time-consuming. Therefore, due to the onerous nature of the task, sensor calibration for a robotic platform is generally undertaken only once, assuming that the calibration parameters will not change over time. This may be a valid assumption for static platforms, but it is often not true for mobile platforms, especially in robotics. In mobile robotics, robots often need to operate in rough terrain, and the assumption that the sensor calibration is not altered during a task is often not true.

Unlike many previously reported methods, here we consider an algorithm for automatic, targetless, extrinsic calibration of a lidar and camera system that is suitable for easy in-field calibration. The proposed algorithm is completely data-driven and uses a MI-based framework to cross-register the intensity and reflectivity information measured by the camera and laser modalities. The outline of the rest of the paper is as follows. In Section 2 we review related work. Section 3 describes the extrinsic laser-camera calibration method. Section 4 presents some calibration results for data collected in indoor and outdoor environments using three different sensor configurations. Section 6 discusses the implications of the laser-camera calibration technique and Section 6 offers some concluding remarks.

2. RELATED WORK

Extrinsic calibration of laser-camera systems is a well-studied problem in computer vision and robotics. The calibration methods reported in the past can be broadly classified into the following two categories: target-based and targetless.

2.1. Target-based

Several methods have been proposed in the past decade that use special calibration targets. One of the most common calibration targets used by researchers, a planar checkerboard pattern, was first used by Zhang (2004) to calibrate a 2D laser scanner and a monocular camera system. He showed that the laser points lying on the checkerboard pattern and the normal of the calibration plane estimated in the camera reference frame provide a geometric constraint on the rigid-body transformation between the camera and laser system. The transformation parameters are estimated by minimizing a nonlinear least squares cost function, formulated by reprojecting the laser points onto the camera image. This was probably the first published method that addressed the problem of extrinsic calibration of lidar/camera sensors in a robotics context. Thereafter, several modifications of Zhang's method have been proposed.

Mei and Rives (2006) reported a similar algorithm for the calibration of a 2D laser range finder and an omnidirectional camera for both visible (i.e., the laser is visible in the camera image also) and invisible lasers. Zhang's method was later extended to calibrate a 3D laser scanner with a camera system (Pandey et al., 2010; Unnikrishnan & Hebert, 2005). Nunnez, Rocha, and Dias (2009) modified Zhang's method to incorporate data from an inertial measurement unit (IMU) into the nonlinear cost function to increase the robustness of the calibration. Mirzaei et al. (2012) provided an analytical solution to the least squares problem by formulating a geometric constraint between the laser points and the plane normal. This analytical solution was further improved by iteratively minimizing the nonlinear least squares cost function. The geometric constraint in planar checkerboard methods requires the estimation of plane normals from camera and laser data. Therefore, the calibration error is correlated to the errors associated with the estimation of these plane normals.

To minimize this error, Zhou and Deng (2012) proposed a new geometric constraint that decouples the estimation of rotation from translation by shifting the origin of the coordinate frame attached to the planar checkerboard target. Recently, Li et al. (2013) proposed an algorithm for extrinsic calibration of a binocular stereo vision system and a 2D lidar. Instead of calibrating each camera of the stereo system independently with the lidar, they proposed an optimal extrinsic calibration method for the combined multi-sensor system based upon 3D reconstruction of the checkerboard target. Although a planar checkerboard target is most common, several other specifically designed calibration targets have also been used in the past. Li et al. (2007) designed a right-angled triangular checkerboard target and used the intersection

points of the laser range finder's slice plane with the edges of the checkerboard to set up the constraint equation. Rodriguez et al. (2008) used a circle-based calibration object to estimate the rigid-body transformation between a multilayer lidar and camera system. Gong et al. (2013) proposed an algorithm to calibrate a 3D lidar and camera system using geometric constraints associated with a trihedral object. Alempijevic et al. (2006) reported a MI-based calibration framework that requires 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 is entirely different due to their requirement of having to track dynamic objects.

2.2. Targetless

The target-based methods require a fiducial object to be concurrently viewed from the lidar and camera sensors, and are therefore not practical for easy *in situ* calibration. Scaramuzza et al. (2007) introduced a technique for the calibration of a 3D laser scanner and omnidirectional camera from natural scenes. They automatically extracted some features from the camera and lidar data and then manually established correspondence between the extracted features. The calibration parameters were then estimated by minimizing the reprojection error for the corresponding points. Recently, Moghadam et al. (2013) proposed a method that exploits the linear features present in a typical indoor environment. The 3D line features extracted from the point cloud and the corresponding 2D line segments extracted from the camera images are used to constrain the rigid-body transformation between the two sensor coordinate frames.

There are also techniques that exploit the statistical dependence of the data measured from the two sensors to obtain a calibration. Boughorbal et al. (2000) proposed a χ^2 test that maximizes the correlation between the sensor data to estimate the calibration parameters. A similar technique was later used by Williams et al. (2004), but their method requires additional techniques to estimate the initial guess of the calibration parameters. Levinson and Thrun (2012) use a series of corresponding laser scans and camera images of arbitrary scenes to automatically estimate the calibration parameters. They use the correlation between the depth discontinuities in laser data and the edges in camera images. A cost function is formulated that captures the strength of the co-observation of depth discontinuity in the laser data and the corresponding edge in the camera image. Recently, Napier et al. (2013) presented a method that calibrates a 2D push broom lidar and a camera system by optimizing a correlation measure between the laser reflectivity and grayscale values from the camera imagery acquired from natural scenes. They do not require the sensors to be mounted such that they have an overlapping field of view, and they compensate for it by observing the same scene at different times from a moving platform. Therefore, they require accurate measurements from an IMU mounted on the moving platform. Recently, Wang et al. (2012) and Taylor and Nieto (2012) have simultaneously proposed similar techniques to our own that use MI as the measure of statistical dependence between the lidar/camera sensor modalities for calibration. Taylor and Nieto (2012) use a MI-based cost function to calibrate a 3D lidar and an omnidirectional camera mounted on a vehicle, and they show that maximizing MI is better than minimizing joint-entropy of the reflectivity and intensity values obtained from these sensors. Similarly, Wang et al. (2012) use normalized mutual information to calibrate a 2D lidar with a hyperspectral camera, and they have shown promising results.

2.3. Our Approach and Contributions

The recent works by Levinson and Thrun (2012) and Napier et al. (2013) are closely related to our own, in the sense that they also propose a fully automatic and targetless method for extrinsic calibration; however, their formulation of the optimization function is quite different. As far as the method is concerned, Wang et al. (2012) and Taylor and Nieto (2012) are the most closely related recent works to our own, although they have either been published at the same time or after Pandey et al. (2012a) (our previous work) and have been used to calibrate specific sensors only. Our previous work explored the idea of using MI-based criteria for automatic calibration of a 3D lidar and an omnidirectional camera. Here, we extend our previous work and show the robustness of the algorithm by performing several different experimental setups using real data obtained from a variety of range/image sensor pairs. In particular, this work builds upon our previous work (Pandey et al., 2012a) to include the following:

- A comprehensive survey of both target-based and targetless methods for calibration of 3D sensors and cameras used in robotics applications.
- A detailed theoretical derivation of the proposed algorithm with implementation details of the kernel density estimate of the probability distribution and the relationship between the joint histogram and the MI-based cost function, which constitutes an important part of the proposed algorithm.
- A comprehensive analysis of the proposed method based on real-world experimental data obtained from three different lidar-camera systems, including *(i) a 3D lidar and omnidirectional camera, (ii) a 3D time-of-flight sensor and monocular camera, and (iii) a 2D lidar and monocular camera*, thereby showing the utility of the proposed algorithm over a wide range of practical applications. Moreover, a comprehensive analysis of the effect of initial conditions and computation time of the algorithm is also included in this paper.

(a) 2D laser scanner (Hokuyo)
(b) 3D laser scanner (Velodyne HDL-64E)
(c) TOF 3D camera (PMD CamCube)
(d) 3D depth camera (Microsoft Kinect)

Figure 3. Various range sensors used in robotics applications.

- A thorough comparison of the proposed algorithm with other state-of-the-art targetless methods (Levinson and Thrun, 2012; Williams et al., 2004) used for calibration in the robotics community.
- An open-source release of the proposed algorithm implemented in C++, used in all experimental results reported here, is available for download from our server at http://robots.engin.umich.edu/SoftwareData/ExtrinsicCalib.

3. METHODOLOGY

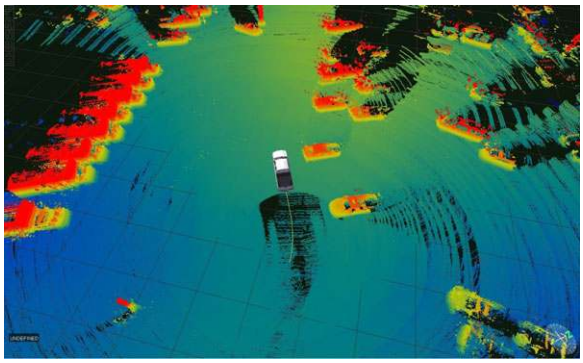
The proposed algorithm is completely data-driven and can be used with any camera, and any range sensor that reports meaningful surface reflectivity values and scene depth information. Various range sensors commonly used in robotics and mapping applications are shown in Figure 3. Most of these sensors report meaningful surface reflectivity values that can be directly used in the proposed algorithm, but for multibeam sensors like the Velodyne (2007), it is important to first perform interbeam calibration of the surface reflectivity values (Levinson & Thrun, 2010). Here, we assume that the reflectivity values are cross-beam calibrated wherever necessary.

In this work, we use the surface reflectivity values reported by the range sensor and the gray-scale intensity values reported by the camera to extrinsically calibrate the two sensor modalities. We claim that under the correct rigid-body transformation, the correlation between the laser reflectivity and the camera intensity is maximized. Our claim is illustrated by a simple experiment shown in Figure 4. Here, we calculate the correlation coefficient for the reflectivity and intensity values for a scan-image pair at different values of the calibration parameter, and we observe a distinct maxima at the true value. Moreover, we observe that the joint histogram of the laser reflectivity and the camera intensity values is least dispersed when calculated under the correct rigid-body transformation.

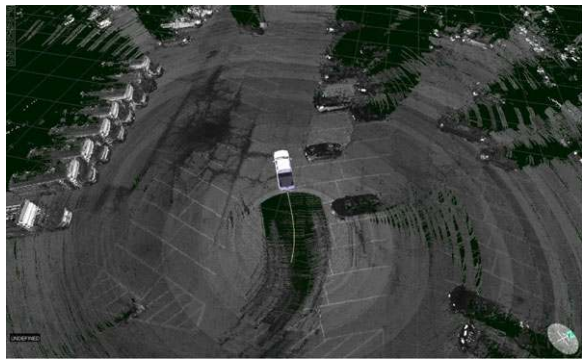
Although scenarios such as Figure 4 do exhibit high correlation between the two modalities, there also exist counterexamples in which the two modalities may not be as strongly correlated, for example infrared absorbing surfaces and shadows. All the lidars that we have used in our experiments emit infrared pulses (Velodyne–905 nm, Hokuyo–870 nm, PMD–950 nm); the reflected light is processed and a reflectivity value based on the amount of energy reflected by the scene is provided to the user. The amount of energy absorbed or reflected back to the lidar depends on the surface properties of the object. Typically, a dark, matte surface absorbs more energy as compared to a light, shiny surface. In our experiments, we use this reflectivity provided by the sensor (after some interbeam calibration) to compute the mutual information with the gray-scale values obtained from the camera. In most of our experiments we observe a reasonable correlation between the lidar reflectivity and the camera intensity because the environment mostly contains objects with either matte or shiny surfaces. If the environment contains colored surfaces that completely absorb infrared, these surfaces will show up as black patches in the lidar reflectivity and will be completely uncorrelated with the corresponding gray-scale values obtained from the camera. Therefore, the mutual-information-based calibration technique might not work well in such situations. However, we are not aware of materials that exhibit such properties (i.e., that completely absorb infrared) and are also found in common indoor/outdoor environments used in robotics applications.

Additionally, in the case of shadows cast in the environment (e.g., see Figure 5), here ambient light plays a critical role in determining the intensity levels of image pixels on the road. As clearly depicted in the image, there are some regions of the road that are covered by object shadows. The gray levels of the image are locally affected by the shadows of occluding objects; however, the corresponding reflectivity values in the laser modality are not because it uses an active lighting principle. Thus, in these types of scenarios, the data between the two sensors might not exhibit 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 are, for the most part, not corrupted by lighting artifacts. In fact, for many practical

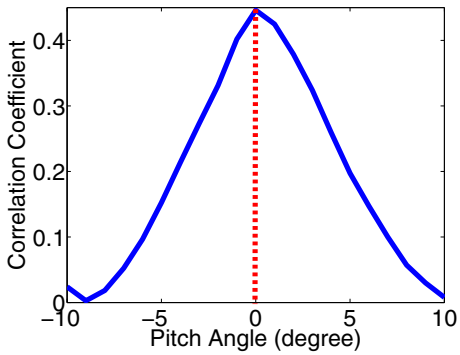
(a) Omnidirectional camera image



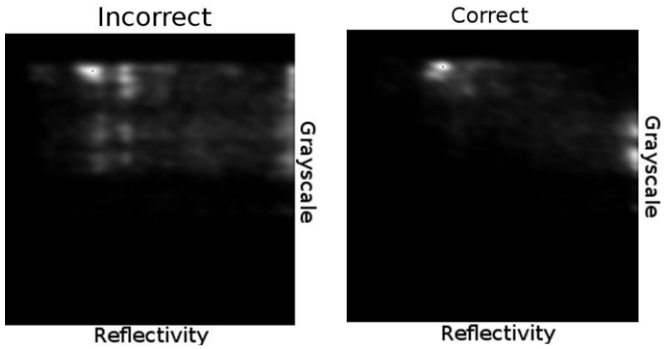
(b) Corresponding lidar colored by height



(c) Corresponding lidar colored by reflectivity



(d) Grayscale/reflectivity correlation



(e) Grayscale/reflectivity joint distribution

Figure 4. Simple experiment illustrating the available correlation between lidar measured surface reflectivity and camera measured image intensity. (a) Image from the Ladybug3 omnidirectional camera. (b) and (c) Depiction of the Velodyne HDL-64E 3D lidar data color-coded by height above ground and by laser reflectivity, respectively. (c) The correlation coefficient for the reflectivity/intensity values as a function of one of the extrinsic calibration parameters, pitch, while keeping all other parameters fixed at their true value. We observe that the correlation coefficient is maximum for the true pitch angle of 0°, denoted by the dashed vertical line. (d) Depiction of the joint histogram of the reflectivity and intensity values when calculated at an incorrect (left) and correct (right) rigid-body transformation. Note that the joint histogram is least dispersed under the correct rigid-body transformation.

indoor/outdoor calibration scenes (e.g., Figure 4), 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 scan views.

3.1. Theory and Background

Mutual information based registration dates back to the early 1990s when Woods et al. (1993) first introduced such a registration method for multimodality images. Their method was based on the assumption that images of the

(a) Image data (b) Lidar data

Figure 5. Counterexample showing that nonuniform lighting can play a critical role in influencing reflectivity/intensity correlation. (a) Ambient lit image with shadows of trees and buildings on the road. (b) Top view of the corresponding lidar reflectivity map, which is unaffected by ambient lighting due to its active lighting principle.

same object taken from different sensors have similar gray-scale values. In a more ideal case, the ratio of gray levels of corresponding points in a particular region of the image should have low variation. Thus, they proposed a method to minimize the average variance of this ratio in order to obtain the registration parameters. Hill et al. (1993) extended this idea to construct a joint histogram of the gray values of the two images, and they showed that the dispersion of the histogram is minimum when the two images are aligned. Soon thereafter, Viola and Wells (1997) and Maes et al. (1997) nearly simultaneously introduced the idea of mutual information for alignment of data captured from two different sensing modalities. The algorithmic developments in MI-based registration were exponential during the late 1990s and early 2000s and very soon became state-of-the-art in the medical image registration field. Researchers widely used the MI framework to focus on specific registration problems in various clinical applications. Within the robotics community, the application of MI has not been as widespread, even though robots today are often equipped with different modality sensors to perceive the environment around them.

The mutual information 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:

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

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

$$\mathrm{H}(X) = -\sum_{x \in X} \mathrm{p}_X(x) \log \mathrm{p}_X(x), \tag{2}$$

$$\mathrm{H}(Y) = -\sum_{y \in Y} \mathrm{p}_Y(y) \log \mathrm{p}_Y(y), \tag{3}$$

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

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

3.2. Mathematical Formulation

Here we consider the laser reflectivity value of a 3D point and the corresponding gray-scale 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 variables, $p(X)$, $p(Y)$, and $p(X, Y)$, can be obtained from the normalized marginal and joint histograms of the reflectivity and gray-scale intensity values of the 3D points co-observed by the lidar and camera. Let $\{\mathbf{P}_i;\, i = 1, 2, \ldots, n\}$ be the set of 3D points whose coordinates are known in the laser reference system, and let $\{X_i;\, i = 1, 2, \ldots, n\}$ 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

$$\tilde{\mathbf{p}}_i = \mathrm{K}[\mathrm{R} \mid t]\tilde{\mathbf{P}}_i, \tag{5}$$

where $(\mathrm{R}, \mathbf{t})$, called the extrinsic parameters, are the orthonormal rotation matrix and translation vector that

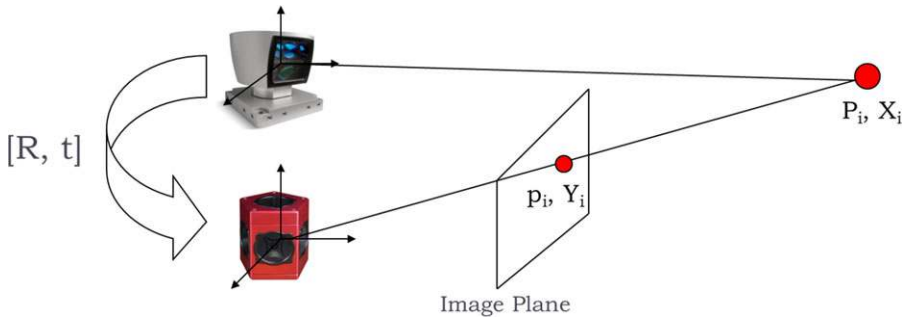


Figure 6. Illustration of the mathematical formulation of MI-based calibration.

relate the laser coordinate system to the camera coordinate system, and K is the camera intrinsics matrix. Here, R is parametrized by the Euler angles $[\phi, \theta, \psi]^\top$, and $\mathbf{t} = [x, y, z]^\top$ is the translation vector. Let $\{Y_i; i = 1, 2, \ldots, n\}$ be the gray-scale intensity value of the image pixel upon which the 3D laser point projects, such that

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

where $Y_i \in [0, 255]$, I is the gray-scale image, and $\mathbf{p}_i$ is the inhomogeneous version of $\tilde{\mathbf{p}}_i$.

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 (Figure 6). The marginal and joint probabilities of the random variables X and Y can be obtained in several different ways. One of the simplest and most commonly used estimators of probability distribution is the maximum likelihood estimator, which is directly obtained from the normalized histogram:

$$\hat{p}(X = k) = \frac{x_k}{n}, \quad k \in [0, 255], \tag{7}$$

where x_k is the observed counts of the intensity value k:

$$x_k = \sum_{i=1}^{n} \mathcal{I}(X_i = k), \tag{8}$$

$$\mathcal{I}(X_i = k) = \begin{cases} 1 & \text{if} \quad X_i = k, \\ 0 & \text{if} \quad X_i \neq k. \end{cases} \tag{9}$$

Although the MLE is easy to compute, generally it has a high mean-squared error (MSE). Therefore, here we use a kernel density estimate (KDE) of the probability distribution, which has been shown to have less MSE as compared to the MLE, and is computed by smoothing the MLE with a symmetric kernel (Scott, 1992):

$$\hat{p}(X = k) = \frac{1}{n} \sum_{i=1}^{n} K_\omega(X - X_i), \quad k \in [0, 255], \tag{10}$$

where $K_\omega(\cdot)$ is a symmetric kernel and ω is the *bandwidth* of the kernel. An illustration of the KDE of the probability distribution of the gray-scale values from the available histogram is shown in Figure 7.

The KDE of the joint distribution of the random variables X and Y is given by

$$\hat{p}(X = k_x, Y = k_y) = \frac{1}{n} \sum_{i=1}^{n} K_\Omega \left(\begin{bmatrix} X \\ Y \end{bmatrix} - \begin{bmatrix} X_i \\ Y_i \end{bmatrix} \right), \quad (k_x, k_y) \in ([0, 255] \times [0, 255]), \tag{11}$$

where $K_\Omega(\cdot)$ is a symmetric kernel and Ω is the the *smoothing* matrix of the kernel. In our experiments we have used a Gaussian kernel, and the smoothing matrix Ω is computed from *Silverman's rule of thumb* (Silverman, 1986):

$$\Omega = 1.06 n^{1/5} \begin{bmatrix} \sigma_X & 0 \\ 0 & \sigma_Y \end{bmatrix}, \tag{12}$$

where σ_X and σ_Y are the standard deviations of the observations of X and Y, respectively.

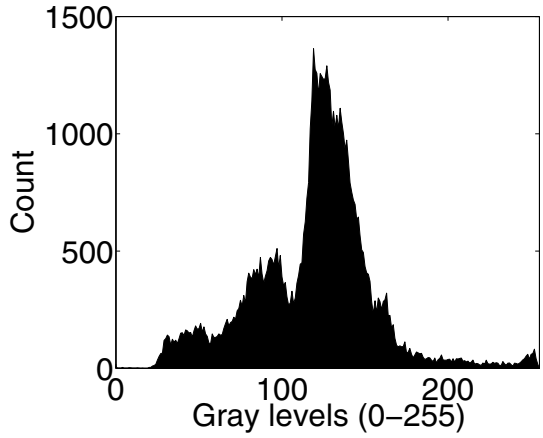
Once we have an estimate of the probability distribution, we can then write the MI of the two random variables as a function of the extrinsic calibration parameters (R, $\mathbf{t}$), thereby formulating an objective function:

$$\hat{\boldsymbol{\Theta}} = \arg\max_{\boldsymbol{\Theta}} \ \mathrm{MI}(X, Y; \boldsymbol{\Theta}), \tag{13}$$

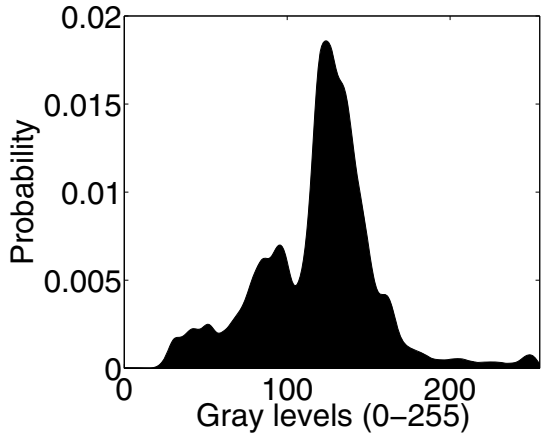
whose maxima occur at the sought-after calibration parameters, $\boldsymbol{\Theta} = [x, y, z, \phi, \theta, \psi]^\top$. KDE provides a smooth and more accurate estimate of the probability distribution, resulting in a distinct optimum in the MI-based objective function, near the correct value of the calibration parameter [see, for example, Figure 7(c)].

3.3. Optimization

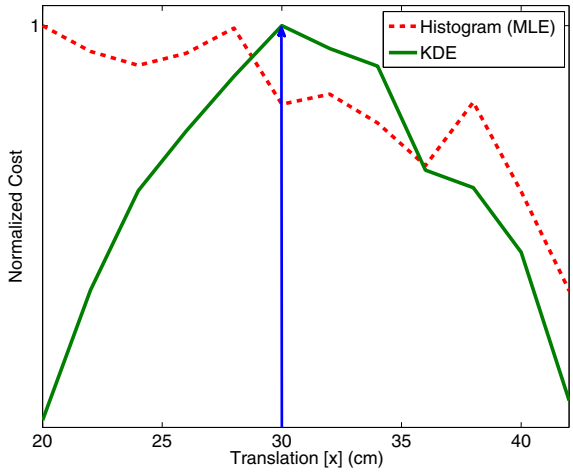
The cost function (13) is maximized at the correct value of the rigid-body transformation parameters. Therefore, any optimization technique that iteratively converges to the global optimum can be used here. Some of the commonly used optimization techniques compute the gradient or Hessian of the cost function (Barzilai & Borwein, 1988; Levenberg, 1944; Marquadrt, 1963; Whittaker & Robinson, 1967). The proposed method does not provide an analytical



(a) Histogram of grayscale intensity values.



(b) KDE of the probability distribution



(c) MI-based cost function for KDE vs. MLE estimate

Figure 7. In (b) we have plotted the KDE of the probability distribution computed from the histogram of sample data. KDE provides a smooth and more accurate estimate of the probability distribution, which in turn affects the cost function. In (c) we have plotted the MI-based cost function (for the same scan-image pair) (i) computed directly from the normalized histogram (red dashed) and (ii) computed from the KDE (green solid). Clearly, when using the KDE we obtain a distinct optimum in the cost function near the correct value (i.e., 30 cm) of the calibration parameter.

derivative of the cost function with respect to the 6-DOF calibration parameters. This is mainly because we cannot write the joint and marginal histograms of the reflectivity and the intensity values as a direct function of the calibration parameters. We first project the 3D point into the image and then establish correspondence between reflectivity and intensity values of the projected point to generate the histograms. Since the histograms are calculated in the image space, the cost function does not involve the calibration parameters in a manner that can be analytically differentiated. Although, the proposed cost function does not have a parametric form, we can still compute the gradient of the cost function numerically and use one of the gradient-descent algorithms to solve the optimization problem. In all of our experiments, we have used the gradient-descent algorithm proposed by Barzilai & Borwein (1988). This method uses 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 (13) is given by

$$\mathbf{G} \equiv \nabla \mathrm{MI}(X, Y; \boldsymbol{\Theta}), \tag{14}$$

then one iteration of the Barzilai & Borwein (1988) method is defined as

$$\boldsymbol{\Theta}_{k+1} = \boldsymbol{\Theta}_k + \gamma_k \frac{\mathbf{G}_k}{\|\mathbf{G}_k\|}, \tag{15}$$

Algorithm 1 Automatic extrinsic calibration by maximization of mutual information

Input: 3D Point cloud $\{\mathbf{P}_i;\ i = 1, \ldots, n\}$, Reflectivity $\{X_i;\ i = 1, \ldots, n\}$, Image {I}, and Initial guess $\{\boldsymbol{\Theta}_0\}$
Output: Estimated parameter $\{\hat{\boldsymbol{\Theta}}\}$
while $\|\boldsymbol{\Theta}_{k+1} - \boldsymbol{\Theta}_k\| > THRESHOLD$ **do**
  $\boldsymbol{\Theta}_k \rightarrow \left[\mathrm{R} \mid \mathbf{t}\right]$
  Initialize the joint histogram: $\mathrm{Hist}(X, Y) = 0$
  for $i = 1 \rightarrow n$ **do**
    $\mathbf{p}_i = \mathrm{K}\left[\mathrm{R} \mid \mathbf{t}\right]\tilde{\mathbf{P}}_i$
    $Y_i = \mathrm{I}(\mathbf{p}_i)$
    Update the joint histogram: $\mathrm{Hist}(X_i, Y_i) = \mathrm{Hist}(X_i, Y_i) + 1$
  end for
  Calculate the kernel density estimate of the joint distribution: $p(X, Y; \boldsymbol{\Theta}_k)$
  Calculate the mutual information: $\mathrm{MI}(X, Y; \boldsymbol{\Theta}_k)$
  Update the current estimate: $\boldsymbol{\Theta}_{k+1} = \boldsymbol{\Theta}_k + \lambda F(\mathrm{MI}(X, Y; \boldsymbol{\Theta}_k))$, where F is either the gradient function or some heuristic and λ is a tuning parameter
end while

where $\boldsymbol{\Theta}_k$ is the optimal solution of (13) at the kth iteration, $\mathbf{G}_k$ is the gradient vector (computed numerically) at $\boldsymbol{\Theta}_k$, $\|\cdot\|$ is the Euclidean norm, and γ_k is the adaptive step size, which is given by

$$\gamma_k = \frac{\mathrm{s}_k^\top \mathrm{s}_k}{\mathrm{s}_k^\top \mathbf{g}_k}, \tag{16}$$

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

Moreover, one can also use heuristic methods (Forrest, 1993; Kirkpatrick, Gelatt, & Vecchi, 1983; Nelder & Mead, 1965) that do not even require the computation of gradients. It should be noted that the proposed cost function has no dependence upon the optimization technique used to solve for the calibration parameters. If the cost function is smooth and exhibits a distinct optimum, then any optimization technique should give the same results. However, we will show in our experiments (Section 4.1.2) that the cost function is not smooth all of the time. In situations in which the cost function is not smooth because of insufficient data, the exhaustive search (computationally expensive) methods are more likely to converge to the correct solution; however, for smooth cost functions with a distinct optimum, the gradient-based or heuristics methods are suitable as they converge to the correct solution within a few steps. The complete MI-based calibration algorithm is shown in Algorithm 1.

3.4. Cramér-Rao Lower Bound of the Estimated Parameter Variance

It is important to know the uncertainty in the estimated calibration parameters in order to use them in any vision or SLAM algorithm. Here we use the Cramér-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 the amount of information that the observations of the random variable Z carry about an unknown parameter α, upon which the probability distribution of Z depends. If the distribution of a random variable Z is given by $p(Z;\alpha)$, then the Fisher information is given by (Lehmann & Casella, 2011)

$$\mathcal{I}(\alpha) = E\left[\left(\frac{\partial}{\partial\alpha}\log p(Z;\alpha)\right)^2\right]. \tag{17}$$

In our case, the joint distribution of the random variables X and Y, as defined in Eq. (11), depends upon the six-dimensional transformation parameter $\boldsymbol{\Theta}$. Therefore, the Fisher information is given by a $[6 \times 6]$ matrix, $\mathcal{I}(\boldsymbol{\Theta})$, whose elements are individually computed as

$$\mathcal{I}(\boldsymbol{\Theta})_{ij} = E\left[\frac{\partial}{\partial\boldsymbol{\Theta}_i}\log p(X, Y; \boldsymbol{\Theta})\frac{\partial}{\partial\boldsymbol{\Theta}_j}\log p(X, Y; \boldsymbol{\Theta})\right] \tag{18}$$

The CRLB is then given by

$$\mathrm{Cov}(\boldsymbol{\Theta}) \geq \mathcal{I}(\boldsymbol{\Theta})^{-1}, \tag{19}$$

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

4. EXPERIMENTS AND RESULTS

This section describes in detail the experiments performed to evaluate the accuracy and robustness of the proposed automatic calibration technique. We present both qualitative and quantitative results with data collected from three different sensor pairs commonly used in robotics applications. The proposed method gives accurate results over the wide range of sensor pairs used.

(a) (b)

Figure 8. (a) The modified Ford F-250 pickup truck with sensor configuration as described in Pandey et al. (2011b). (b) The Velodyne HDL-64E 3D laser scanner (Velodyne, 2007) and the Point Grey Ladybug3 omnidirectional camera (Point Grey Research Inc., 2009) are mounted on the roof of the vehicle.

4.1. 3D Laser Scanner and Omnidirectional Camera

In the first set of experiments, we present calibration results from a Velodyne HDL-64E 3D laser scanner (Velodyne, 2007) and a Point Grey Ladybug3 omnidirectional camera system (Point Grey Research Inc., 2009) mounted on the roof of a vehicle (Figure 8). In this work, we pre-calibrated the reflectivity values of the Velodyne laser scanner using the algorithm reported by Levinson & Thrun (2010), and we used the manufacturer provided intrinsic calibration parameters (focal length, camera center, distortion coefficients of the lens) for the omnidirectional camera. In all of our experiments in this section, *scan* refers to a single 360° field-of-view 3D point cloud and its time-corresponding camera imagery.

4.1.1. Calibration Performance Using a Single Scan

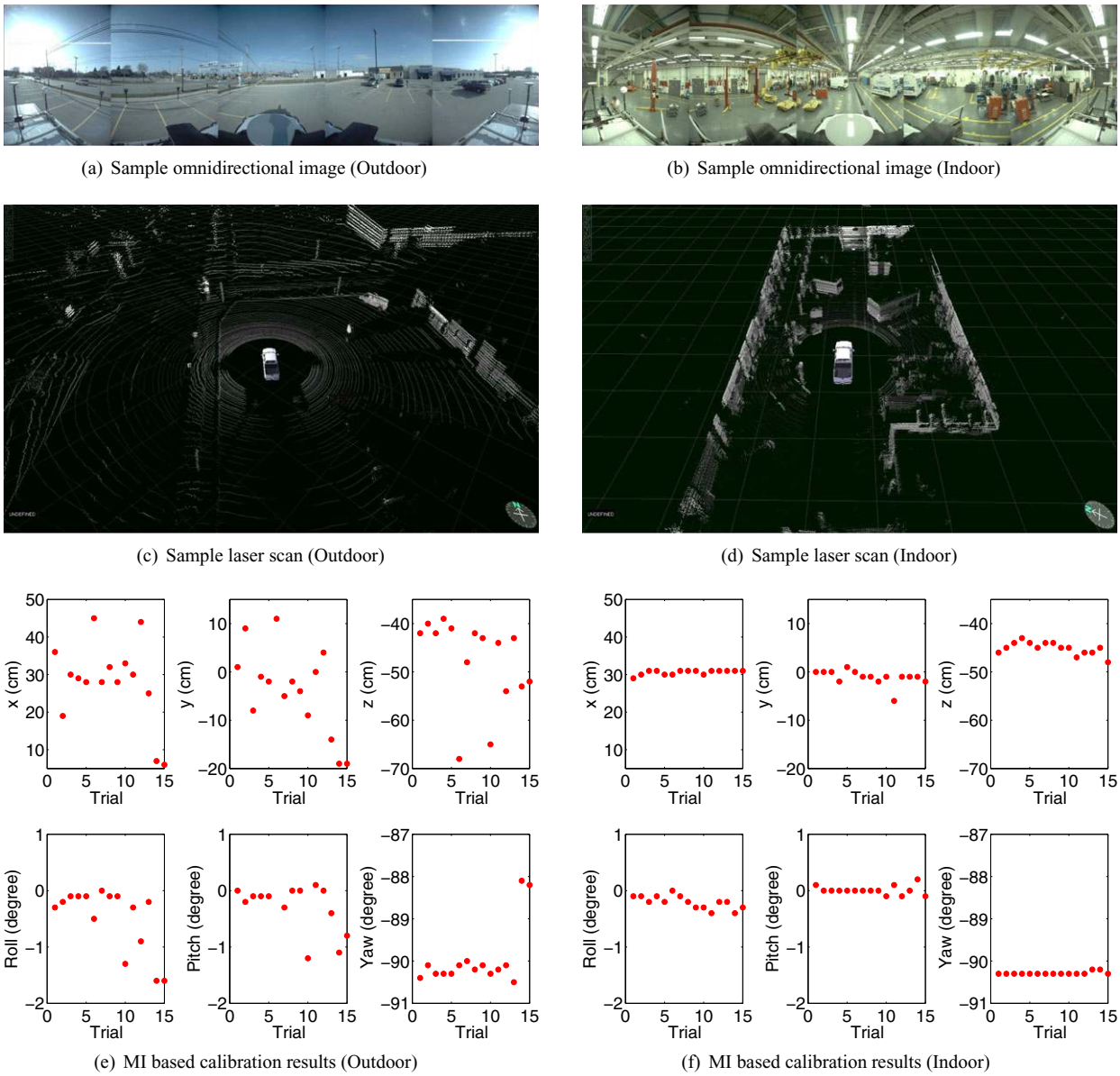
In this experiment, we show that the quality of the *in situ* calibration performance is dependent upon the environment in which the scans are collected. We collected several datasets in both indoor and outdoor settings. The indoor dataset was collected inside a large garage, and it exhibited many near-field objects such as walls and other vehicles. In contrast, the outdoor dataset includes lighting artifacts (Figure 4), moving objects, and most of the structure lying in the far-field. In Figures 9(e) and 9(f), 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. This is not surprising as the outdoor dataset is more likely to be corrupted with lighting artifacts and dynamic objects; however, we observe that the error fluctuation in translation parameters is higher as compared to rotational parameters. We attribute this asymmetric error behavior to the presence of only far-field 3D points in the outdoor dataset, rendering the cost function less sensitive to the translational calibration 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 vanishing points) is not affected by the translational component of the camera projection matrix (Hartley & Zisserman, 2000):

$$\tilde{\mathbf{p}} = \mathrm{K}\Big[\mathrm{R} \mid \mathbf{t}\Big]\begin{bmatrix}\tilde{x}\\ \tilde{y}\\ \tilde{z}\\ 0\end{bmatrix} = \mathrm{KR}\begin{bmatrix}\tilde{x}\\ \tilde{y}\\ \tilde{z}\end{bmatrix}. \tag{20}$$

We should expect then that scans that only contain 3D points far off in the distance (i.e., the outdoor dataset) will have poor observability of the calibration parameters, as opposed to scans that contain many nearby 3D points (i.e., the indoor dataset), as seen in Figure 9. Therefore, if we intend to perform MI-based calibration from a single scan-image pair, we should use data collected with this effect in mind.

4.1.2. Calibration Performance Using Multiple Scans

In the previous section, we showed that it is necessary to have near-field objects, no lighting artifacts, and no moving objects in the scans in order to robustly estimate the calibration parameters from a single scan; however, this might not always be practical—depending upon the operational environment. In this experiment, we demonstrate improved calibration convergence by simply aggregating multiple scans into a single batch optimization process (Figure 10). It should be noted that the reflectivity from lidar and



(a) Sample omnidirectional image (Outdoor)

(b) Sample omnidirectional image (Indoor)

(c) Sample laser scan (Outdoor)

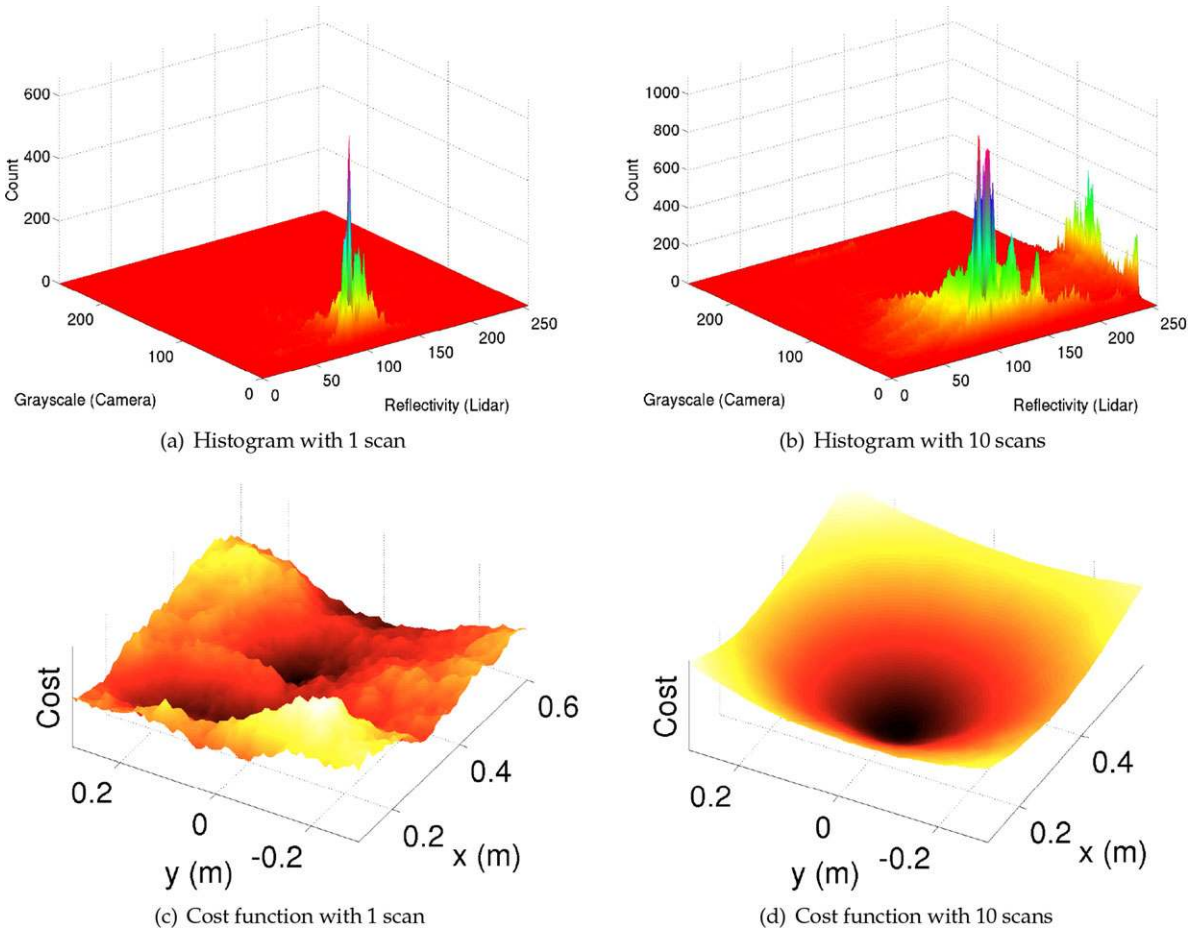
(d) Sample laser scan (Indoor)

(e) MI based calibration results (Outdoor)

(f) MI based calibration results (Indoor)

Figure 9. 3D laser and omnidirectional camera single-view calibration results for outdoor and indoor datasets. The variance in the estimated parameters (especially translation) is significantly large in the case of the outdoor dataset due to poor observability as noted in the text. Each point on the abscissa in (e)–(f) corresponds to a single scan trial.

gray-scale intensity from the camera is quantized between [0, 255], resulting in a large joint histogram ($256 \times 256 = 65,536$ bins) that needs to be estimated. The number of 3D points or observations (X_i, Y_i) of these random variables obtained from a single scan when using Velodyne data is typically of the order of 80,000 points. Therefore, if we use a single scan-image pair, the joint histogram is largely undersampled [Figure 10(a)] because only about 80,000 observations are used to populate a histogram of 65,536 bins. However, if we use more data (i.e., scan-image pairs from multiple locations) to generate the joint histogram, they fill in the unobserved sections of the histogram [Figure 10(b)]. This results in a better estimate of the joint and marginal probability distributions of the random variables, which in turn improves the MI estimate and increases the smoothness of the cost function [Figure 10(d)]. The smooth cost function now exhibits a distinct optimum near the correct calibration parameters. We can therefore use any gradient descent algorithm



(a) Histogram with 1 scan

(b) Histogram with 10 scans

(c) Cost function with 1 scan

(d) Cost function with 10 scans

Figure 10. The top panel shows the joint histogram of lidar reflectivity and camera intensity values. We get a better estimate of the joint histogram (fill-in of unobserved sections) as the number of scan-image pairs is increased. The bottom panel shows the MI cost-function surface versus translation parameters x and y. Note the distinct optimum and smoothness of the cost surface 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 to see.

to quickly converge to the global optimum of this cost function.

Figure 11 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 is increased by just a few. Here, the red plot shows the sample standard deviation (σ) of the calibration parameters computed over 1,000 trials, where in each trial we randomly sampled $\{N = 5, 10, \ldots, 40\}$ 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 20–40 scans, we can achieve very accurate performance. Moreover, we see that the sample variance asymptotically approaches the CRLB as the number of scans used increases, indicating that this is an efficient estimator. In this experiment, we took static snapshots of the laser scan and the camera image to avoid any errors due to motion of the vehicle. Although using the static snapshot is the best way to acquire data for calibration, if we have access to a good IMU mounted on the vehicle, the calibration process can be made even more user-friendly. In that case, we can motion-compensate the scan data using the IMU and then use them in the proposed calibration method. This allows for easy online calibration of the sensors without the need for acquiring static snapshots. We found that the calibration parameters obtained from the motion-compensated scans (using a good IMU) are close to those obtained from the static scans (Table I).

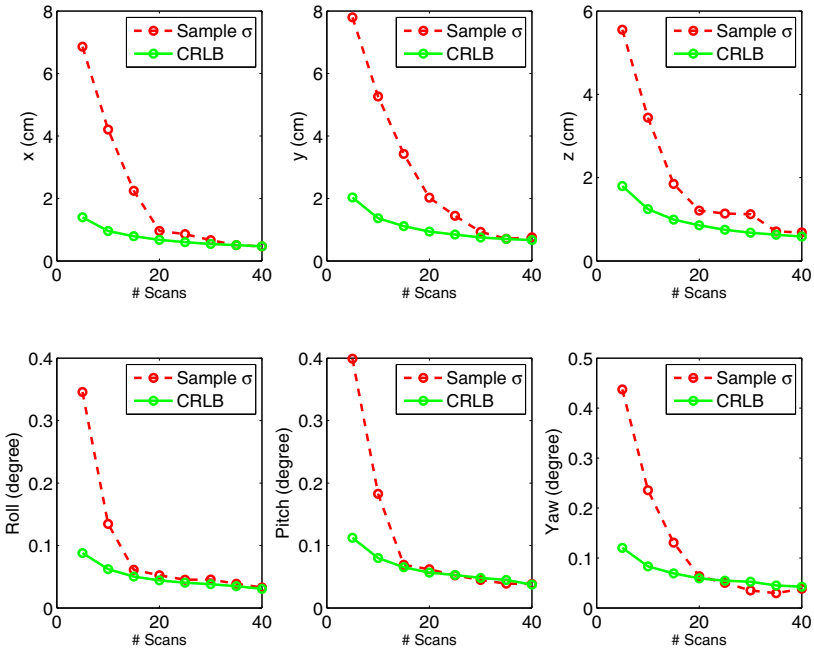


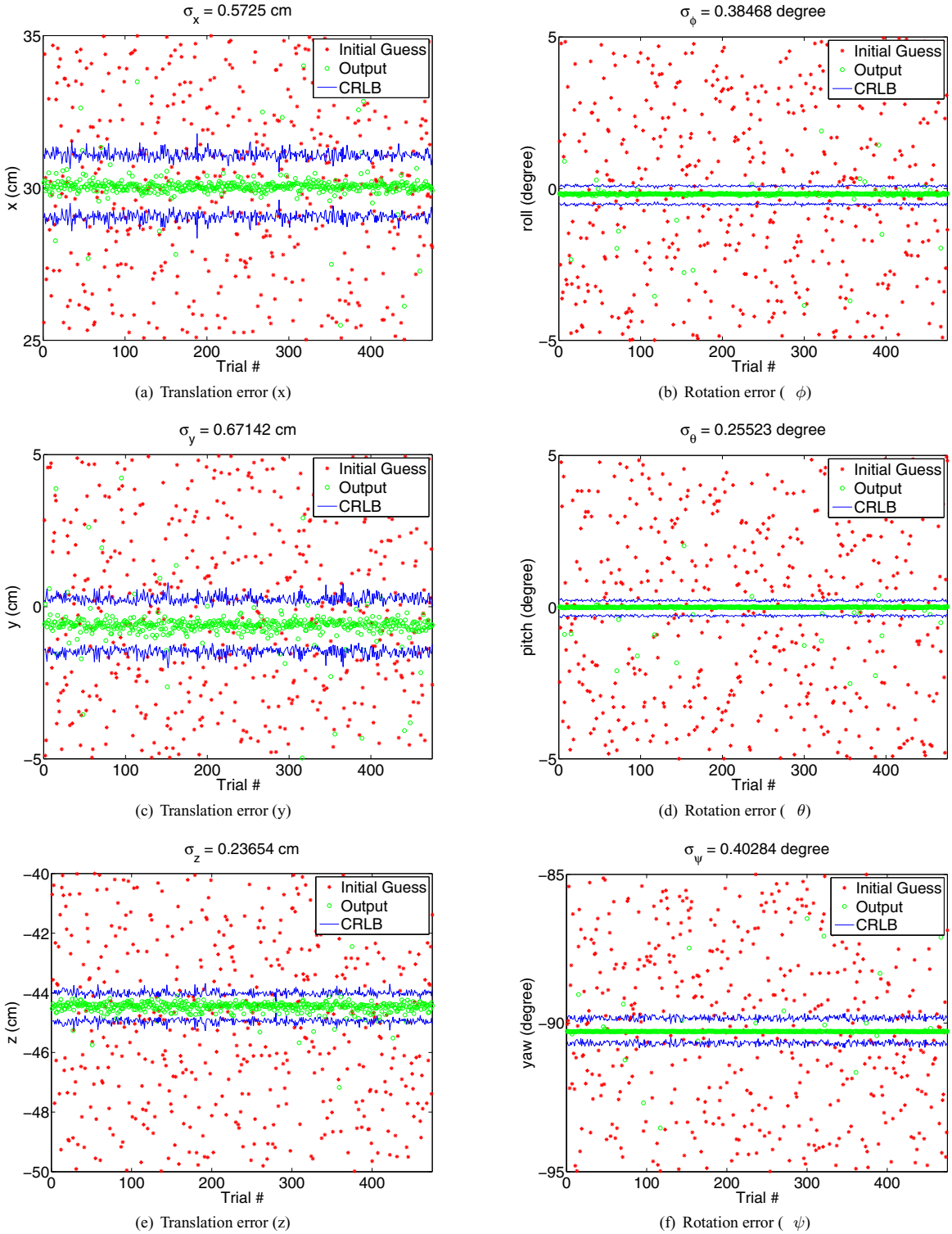
Figure 11. 3D laser and omnidirectional camera multiview calibration results. Here we use all five horizontal images from the Ladybug3 omnidirectional camera during the calibration. Plotted is the uncertainty of the recovered calibration parameters versus the number of scans used. The red (dashed line) plot shows the sample-based standard deviation (σ) of the estimated calibration parameters calculated over 1,000 trials. The green (solid line) 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 I. Comparison of calibration parameters estimated by the proposed method with static scans, the proposed method with motion-compensated scans, feature alignment as reported in Levinson and Thrun (2012) for 40 and 100 scan pairs, a χ^2 test as reported in Williams et al. (2004), and a checkerboard target pattern as reported in Pandey et al. (2010).

Method	Data	x (cm)	y (cm)	z (cm)	Roll (deg)	Pitch (deg)	Yaw (deg)
Proposed method	40 Scans	30.5	−0.5	−44.6	−0.15	0.00	−90.27
w/motion compensated	40 Scans	33.6	−0.7	−41.6	−0.20	−0.06	−90.14
Levinson and Thrun (2012)	40 Scans	30.0	0.7	−40.7	−0.37	−0.24	−89.72
	100 Scans	31.6	−0.3	−41.7	−0.05	0.05	−90.12
Williams et al. (2004)	40 Scans	29.8	0.0	−43.4	−0.15	0.00	−90.32
Pandey et al. (2010)	14 Planes	34.0	1.0	−41.6	0.01	−0.03	−90.25

4.1.3. Calibration Performance with Different Initial Guesses

In this experiment, we show the robustness of the proposed algorithm over the initial guess of the calibration parameters. As described in Algorithm 1, the proposed algorithm requires an initial guess of the calibration parameters, which is generally obtained by manually measuring the distances and angles between the two sensors. Typically, the error in this measurement is of the order of 10 cm for translation parameters and 10° for rotation parameters. So, here we performed 500 independent trials with a random initial guess (within the measurement errors), and we observed that the algorithm converges to the correct calibration parameters (Figure 12). In this experiment, we used 20 randomly sampled scan-image pairs from our indoor and outdoor dataset. We observe that the standard deviation of the estimated translation parameters over these 500 trials is less than 0.7 cm and the standard deviation of the rotation parameters is less than 0.5°. Therefore, this experiment clearly depicts the robustness of the proposed algorithm over a wide range of initial guesses of the calibration parameters that is within the acceptable range of manual errors.



(a) Translation error (x)

(b) Rotation error (ϕ)

(c) Translation error (y)

(d) Rotation error (θ)

(e) Translation error (z)

(f) Rotation error (ψ)

Figure 12. Calibration performance for different initial conditions with 20 scan-image pairs. Here we perform 500 independent trials with random initial guess. The initial guess is marked in red, the output of the proposed calibration algorithm is marked in green, and the CRLB of the standard deviation of the estimated parameters is shown in blue.

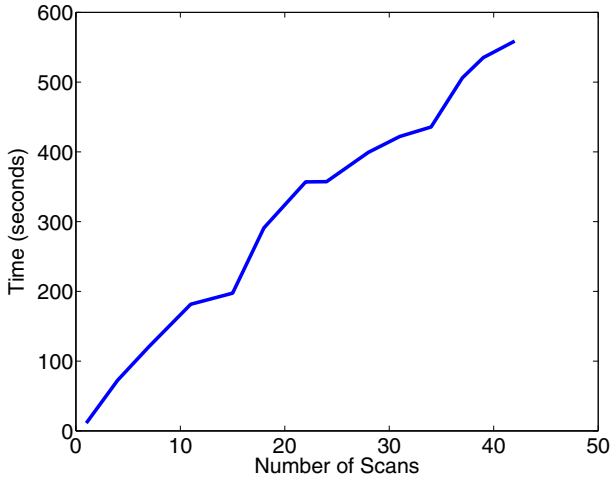


Figure 13. Computation time as a function of the number of scans used for calibration. Computation time increases as the number of data points are increased (one scan contains approximately 80,000–100,000 3D points). More data result in better calibration performance, so there is a tradeoff between computation time and the robustness of the algorithm.

4.1.4. Computation Time Analysis

In this experiment, we analyzed the computational complexity of the proposed algorithm. In Section 4.1.2, we showed that as we increase the number of scans, from different viewpoints, the calibration performance increases. However, the increase in the number of scans also increases the computation time of the algorithm. Since the computational complexity of the algorithm is $O(n + m^2)$, where n is the number of 3D points used and m is the number of quantization bins of the random variables X and Y, if the number of bins is fixed (here 256), then the computation time increases linearly with the increase in the number of 3D points or scans. Figure 13 shows a plot of computation time as a function of the number of scans used with a simple gradient descent algorithm (Barzilai & Borwein, 1988) as the optimization method. We observe that the computation time (on a standard laptop with Intel Core i7-2670QM CPU @ 2.20 GHz) when the algorithm uses 20 scan-image pairs is of the order of 5 min. There is a clear tradeoff between the computation time and the robustness of the algorithm as the increase in the number of scans makes the algorithm more robust but it also increases the computation time. Since calibration is typically an offline task, there is no need for the algorithm to be real-time; however, we also do not want to wait for very long to obtain the results. Therefore, an optimal value of the number of scans should be chosen depending upon the application. In our experiments, we observed that 20–40 scans provide a good calibration result (Sections 4.1.2 and 4.1.3) within 5–10 min, respectively, which we believe is acceptable for practical in-field operations of robots. We would like to point out that the current implementation of the algorithm (used in our experiments) is unoptimized as we compute the joint histogram in a serialized fashion iterating through every point in the scans; however, the computation of the joint histogram from the 3D points can be easily parallelizable and is readily multithread or graphics processing unit (GPU) applicable, which could significantly improve upon the times reported here.

4.1.5. Comparison with Other Calibration Methods

We performed the following three experiments to quantitatively benchmark results from our proposed method against other published methods:

1. Comparison with Williams et al. (2004): In this experiment, we replace the MI criteria with 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:

$$\chi^2(X, Y; \boldsymbol{\Theta}) = \sum_{x \in X,\, y \in Y} \frac{\left[p(x, y; \boldsymbol{\Theta}) - p(x; \boldsymbol{\Theta}) p(y; \boldsymbol{\Theta})\right]^2}{p(x; \boldsymbol{\Theta}) p(y; \boldsymbol{\Theta})}. \quad (21)$$

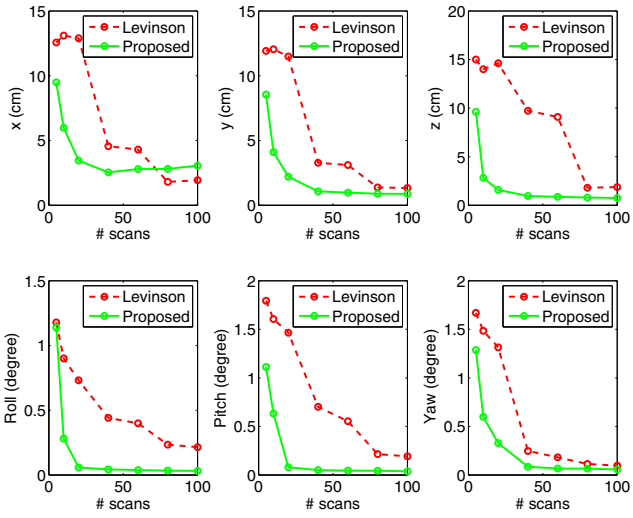
We can therefore modify the cost function given in Eq. (13) to

$$\boldsymbol{\Theta} = \arg\max_{\boldsymbol{\Theta}} \chi^2(X, Y; \boldsymbol{\Theta}). \quad (22)$$

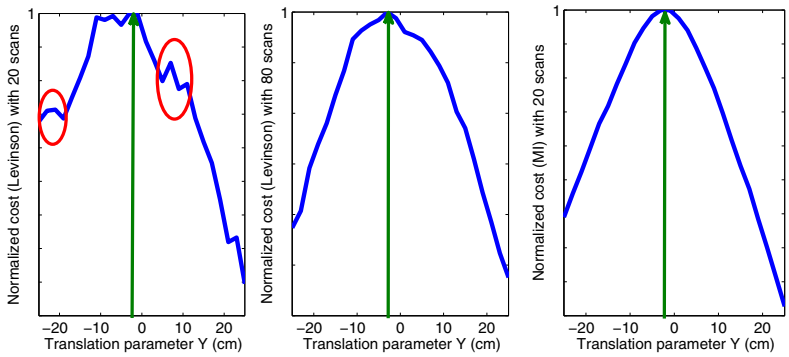
A comparison of the calibration results obtained from the χ^2 test (22) and the MI cost function (13) using 40 scan-image pairs is shown in Table I. We see that the results obtained from the χ^2 statistics are similar to those obtained from the MI criteria. This is mainly because the χ^2 statistics and MI are equivalent and essentially capture the amount of correlation between the two random variables (McDonald, 2009). We use MI as the measure of correlation between the random variables mainly because it is well studied and has been successfully used in practical applications (e.g., medical image registration). Moreover, several researchers have developed methods that robustly and efficiently estimate MI from the sample data [e.g., Chao and Shen (2003), Lin and Medioni (2008), and Hausser and Strimmer (2009)], which can be readily used within the proposed framework.

2. Comparison with Levinson and Thrun (2012): Levinson and Thrun (2012) proposed an automatic calibration technique that uses correlation between depth discontinuities in the laser data and their projected edges in the corresponding camera images. In this experiment, we replace our MI-based cost function with the criteria proposed by Levinson and Thrun:

$$LC(X, Y; \boldsymbol{\Theta}) = \sum_{f=1}^{N} \sum_{p=1}^{|X^f|} X_p^f \cdot D_{i,j}^f, \quad (23)$$



(a) Standard deviation of estimated parameters



(b) Normalized cost for Levinson with 20, Levinson with 80, and MI-based with 20 scans

Figure 14. Comparison with Levinson and Thrun (2012). (a) Here we plot the uncertainty of the recovered calibration parameters versus the number of scans used. The red (solid line) plot shows the sample-based standard deviation (σ) of the estimated calibration parameters calculated over 1,000 trials using Levinson's method (Levinson and Thrun, 2012). The green (dashed line) plot shows the sample-based standard deviation of the estimated parameters using our proposed method. Each point on the abscissa corresponds to the number of aggregated scans used per trial. Clearly the proposed method converges to a good solution with significantly fewer scans. (b) Here we plot the cost computed from the two techniques as a function of one of the calibration parameters (i.e., translation in y). The leftmost plot shows the cost computed by Levinson's method with 20 scan-image pairs from an outdoor dataset (with motion-compensated 3D points). The rightmost plot shows the MI-based cost computed for the same set of scan-image pairs. Clearly the MI-based cost function (computed from 20 scans) is smooth and exhibits a distinct optimum near the correct calibration parameter. Levinson's cost function (computed from the same 20 scans), on the other hand, is rough and shows local optima (marked in red on the leftmost plot); however, Levinson's cost function becomes smooth as we increase the number of scans (center plot). Levinson's cost computed with 80 scan-image pairs is smooth and exhibits a distinct optimum. The correct value of translation parameter is marked with a green arrow, and all the plots show optima at that location.

where $LC(\cdot)$ is Levinson's criterion for N scan image pairs, X_p^f is the depth discontinuity at the pth point in scan f, and $D_{i,j}^f$ is the edge strength at projection of 3D point p onto the corresponding image f. The modified cost function can be written as

$$\boldsymbol{\Theta} = \arg\max_{\boldsymbol{\Theta}} LC(X, Y; \boldsymbol{\Theta}). \tag{24}$$

Figure 14 shows a comparison of the proposed method with Levinson's method. In this experiment, we used motion-compensated scans captured in an outdoor urban environment (Pandey et al., 2011b) to estimate the rigid-body transformation from both methods. In Levinson's method, only points corresponding to the edges of the surfaces are used, discarding a large amount of points corresponding to the ground plane and other flat surfaces present

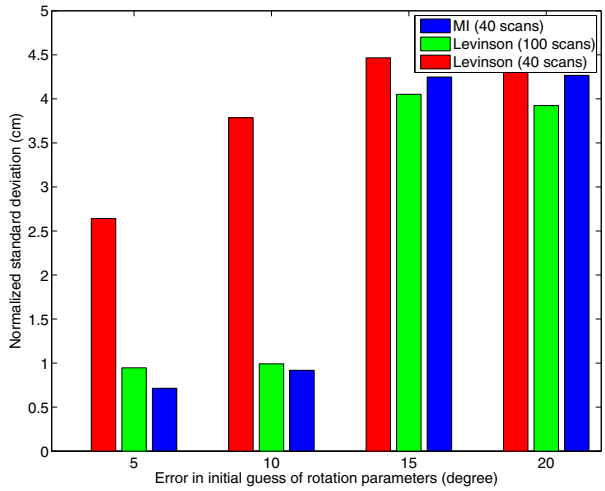
in the environment—therefore, it requires a relatively large number of scans and a structured calibration environment. Although the plots show that for both methods the sample-based standard deviation of the estimated calibration parameters decreases as the number of scans is increased, the proposed method gives good calibration results with only 20 scans, whereas Levinsons method requires nearly 100 scans to reach the same precision level. Unlike Levinson's method, our proposed method is whole-image based and uses *all* of the overlapping laser-image data. This allows our method to produce good calibration results with fewer scans even if the calibration environment is largely devoid of any linear depth discontinuities—the only criterion being that the scene have some distinctive reflectivity/intensity texture (e.g., a parking lot with painted parking stalls as in Figure 4). It should be noted that both methods are equally affected by the lighting artifacts and noise due to moving objects in the scene. However, the noise due to these artifacts is reduced by using more data to compute the cost function in both of the methods. In Figure 14(b), we have plotted the cost computed from both of the methods, and we observe that Levinson's cost function exhibits several local optima when computed with fewer scans. This is also the reason why we observe high variance in the estimated parameters when we use fewer scans in Levinson's method, as the gradient-based optimization technique often gets stuck in these local optima.

Fundamentally, both methods are quite similar as they use the joint statistics of data to compute the calibration parameters. Since Levinson's method does not use the reflectivity value from the lidar and only uses the depth information, this method can be easily used with sensors that do not provide the reflectivity information. The proposed method, on the other hand, only works with sensors that also provide reflectivity information along with the depth (or 3D) information.

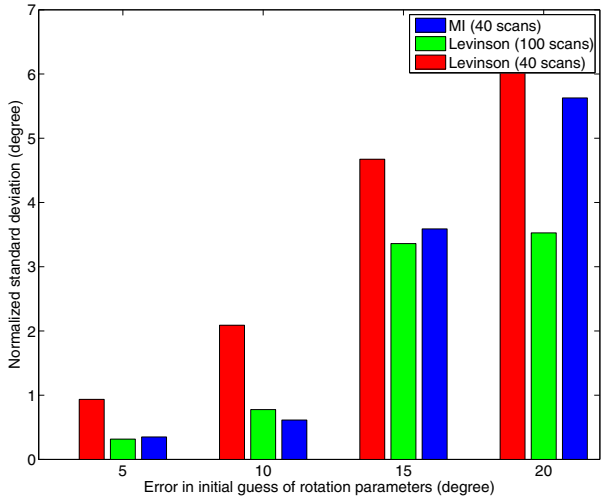
3. Comparison with Pandey et al. (2010): Pandey et al. (2010) proposed a method that requires a planar checkerboard pattern to be observed simultaneously from the laser scanner and the camera system. The normal of the planar surface and 3D points lying on the surface constrain the relative transformation between the laser scanner and the omnidirectional camera system. These constraints are used to solve for the extrinsic calibration parameters within a nonlinear optimization framework. The 3D points lying on the planar surfaces are manually extracted from the 3D point cloud. The normals of these planar surfaces in the camera reference frame are also calculated by manually clicking the corners of the checkerboard pattern in the image. 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 we found that they are very close (Table I). The reprojection of 3D points onto the image using results obtained from these methods looks very similar visually. Therefore, in the absence of ground truth, it is difficult to say which result is more accurate. The proposed method, however, is definitely much faster and easier as it does not involve any manual intervention.

4.1.6. Convergence basin analysis

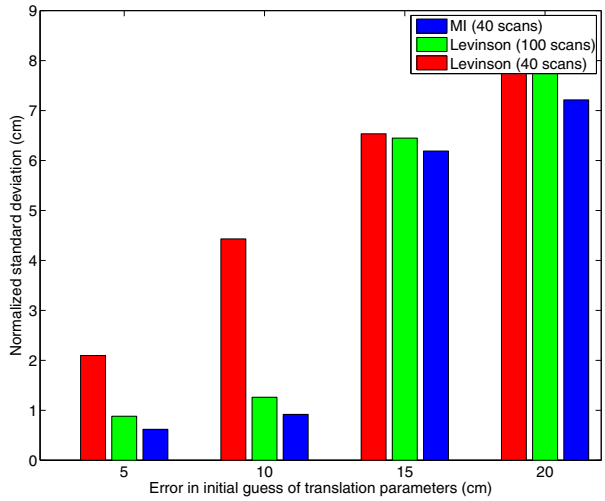
In this experiment, we analyze the convergence basin of the proposed algorithm and compare it with Levinson's algorithm (Levinson and Thrun, 2012). We randomly selected N scan-image pairs from our outdoor dataset and used them to estimate the calibration parameters for several different values of errors in the initial guess provided to the two algorithms. We considered eight different sets of errors in the initial guess ranging from (5 cm, 5°) to (20 cm, 20°). We performed 500 independent trials for each of these errors in the initial guess and plotted the normalized standard deviation of estimated parameters. If the standard deviation of calibration parameters is given by $[\sigma_x, \sigma_y, \sigma_z, \sigma_\phi, \sigma_\theta, \sigma_\psi]$, then the normalized standard deviation of translation and rotation parameters can be written as $\sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}$ and $\sqrt{\sigma_\phi^2 + \sigma_\theta^2 + \sigma_\psi^2}$, respectively. In Figure 15, we have plotted the normalized standard deviation of the estimated calibration parameters computed from 500 independent trials with random initial guesses. The initial guess in each trial is a sample from a uniform distribution with mean equal to the true value of the calibration parameter and standard deviation equal to the error in the initial guess. Figures 15(a) and 15(b) show the standard deviation of estimated parameters when the error in the initial guess of translation parameters is 10 cm (±5 cm) and the error in the initial guess of rotation parameters is changed from 5° to 20°. Figures 15(c) and 15(d) show the standard deviation of estimated parameters when the error in the initial guess of rotation parameters is 10° (±5°) and the error in the initial guess of translation parameters is changed from 5 to 20 cm. We observe that the proposed algorithm provides good results with 40 scans when the error in the initial guess of translation and rotation parameters is within 10 cm and 10°, respectively. On the other hand, Levinson's method with 40 scans fails to converge to the correct solution even for low values of error in the initial guess (e.g., 5 cm, 5°). However, as we increase the number of scans in Levinson's algorithm to 100, it gives results similar to the proposed algorithm, with a convergence basin of 10 cm in translation and 10° in rotation. It should be noted that the straightforward gradient descent optimization technique used in this experiment makes no provisions to avoid local optima, and that while there exist more sophisticated stochastic optimization techniques that can be used to avoid such local optima (Forrest, 1993; Kirkpatrick et al., 1983; Wenzel & Hamacher, 1999), they are not employed here. Moreover, the convergence basin (especially for translation parameters) is also dependent upon the environment in which the data are collected. In this experiment, we have only used motion-compensated scans from



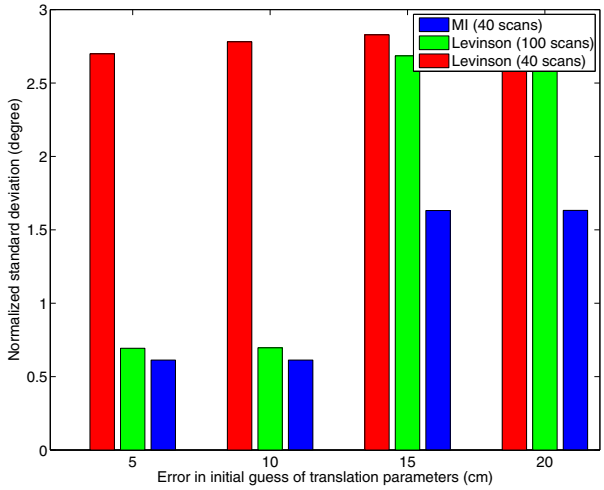
(a) Standard deviation of estimated translation when error in initial guess of translation parameters is 10 cm and error in rotation parameters is changed from 5° to 20°



(b) Standard deviation of estimated rotation when error in initial guess of translation parameters is 10 cm and error in rotation parameters is changed from 5° to 20°

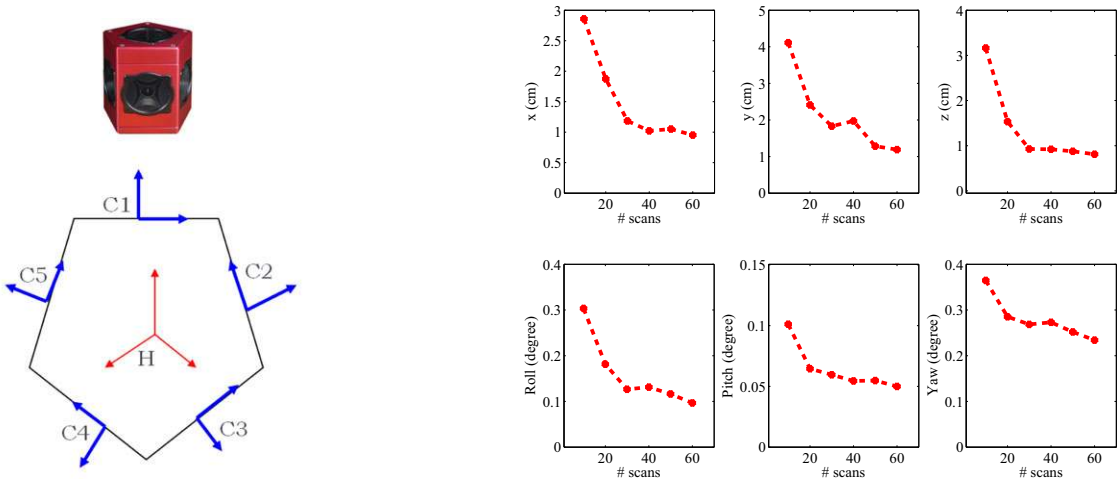


(c) Standard deviation of estimated translation when error in initial guess of rotation parameters is 10°and error in translation parameters is changed from 5 cm to 20 cm



(d) Standard deviation of estimated rotation when error in initial guess of rotation parameters is 10° and error in translation parameters is changed from 5 cm to 20 cm

Figure 15. Convergence basin analysis. Here we have plotted the normalized standard deviation of estimated calibration parameters computed from 500 independent trials with random initial guess. The initial guess in each trial is a sample from the uniform distribution with mean equal to the true value of calibration parameter and standard deviation equal to the error in initial guess. Parts (a) and (b) show the standard deviation of estimated parameters when the error in initial guess of translation parameters is 10 cm and the error in initial guess of rotation parameters is changed from 5° to 20°. Parts (c) and (d) show the standard deviation of estimated parameters when the error in initial guess of rotation parameters is 10° and the error in initial guess of translation parameters is changed from 5 to 20 cm. Red and green bars show the standard deviation of calibration parameters obtained from Levinson's method with 40 and 100 scans, respectively. The standard deviation of calibration parameters from the proposed method using 40 scans is shown in blue. We observe that we obtain good results when the error in initial guess of translation and rotation parameters is within 10 cm and 10°, respectively.



(a) LB3 camera layout

(b) Calibration error w.r.t. manufacturer provided values

Figure 16. Comparison with manufacturer ground-truth. (a) A depiction of the coordinate frames corresponding to each camera (c_i) and camera head (H) of the Ladybug3 omnidirectional camera system. (a) Plotted are the mean absolute error in the relative-pose calibration parameters for the two side-looking cameras (c_2 and c_5), i.e., $|\mathrm{X}_{c_2c_5} - \hat{\mathrm{X}}_{c_2c_5}|$, versus the number of scans used to estimate these parameters. The mean is calculated over 100 trials of sampling N scans per trial $\{N = 10, 20, \ldots, 60\}$. We see that the error decreases as the number of scans is increased.

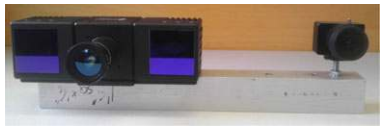
an outdoor dataset, and we observe that the translation parameters are more sensitive to errors in the initial guess. We have already discussed this effect of faraway 3D points on the translation parameters in our previous experiment (Section 4.1.1). Therefore, we can further increase the convergence basin of the proposed algorithm by using indoor scan data and a more sophisticated stochastic optimization technique.

4.1.7. Comparison with Available Ground-truth

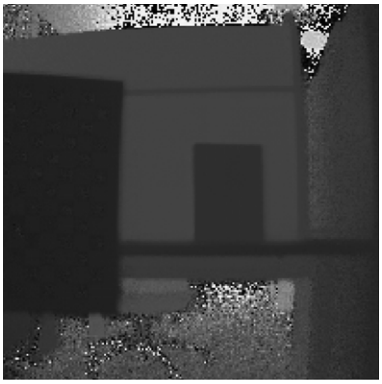
The omnidirectional camera used in our experiments is precalibrated from the manufacturer. It has six 2-Megapixel cameras, with five cameras positioned in a horizontal ring and one positioned vertically, such that the rigid-body transformation of each camera with respect to a common coordinate frame, called the camera head (H), is well-known (Point Grey Research Inc., 2009). Here, X_{Hc_i} is the Smith, Self, and Cheeseman (1988) coordinate frame notation, and it represents the 6-DOF pose of the ith camera (c_i) with respect to the camera head (H). Since we know X_{Hc_i} from the manufacturer, we can calculate the pose of the ith camera with respect to the jth camera as

$$\mathrm{X}_{c_ic_j} = \ominus \mathrm{X}_{Hc_i} \oplus \mathrm{X}_{Hc_j}, \quad \{i \neq j\}. \tag{25}$$

In the previous experiments, we used all five 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 ($\mathrm{X}_{\ell c_i}$). This allows us to calculate $\hat{\mathrm{X}}_{c_ic_j}$ from the estimated calibration parameters $\hat{\mathrm{X}}_{\ell c_i}$ and $\hat{\mathrm{X}}_{\ell c_j}$. Thus, we can compare the true value of $\mathrm{X}_{c_ic_j}$ (from the manufacturer data) with the estimated value $\hat{\mathrm{X}}_{c_ic_j}$. Figure 16 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. Ideally, this error should asymptotically approach the expected value of the error (i.e., $E[|\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}|] \to 0$), however we observe some residual bias. The primary reason for the residual bias is the assumption that the intrinsic parameters of the lidar and camera obtained from the manufacturers are correct, which is not necessarily true. The intrinsic parameters for each laser beam (a total of 64 lasers) of the Velodyne lidar includes (i) elevation angles, (ii) range bias, (iii) intensity calibration, and (iv) rotation angle. Also, for the Ladybug3 omnidirectional camera, the intrinsic parameter of each lens includes (i) focal length, (ii) camera center, (iii) lens distortion parameters, and (iv) relative transform of each camera with respect to the camera head (X_{Hc_i}). Therefore, with so many other parameters that are used to compute the cost function, it is difficult to identify the actual source of the residual bias that we observe in this experiment. It should be noted that in this experiment we used only a single camera as opposed to all five cameras of the omnidirectional camera system, thereby reducing the amount of data used in each trial by 1/5. It is our conjecture that with additional trials, a statistically significant validation of unbiasedness could be achieved.



(a) A 3D TOF camera and a monocular camera mounted on a rigid bar



(b) Depth map from TOF camera

(c) Intensity map from TOF camera

(d) Image from monocular camera

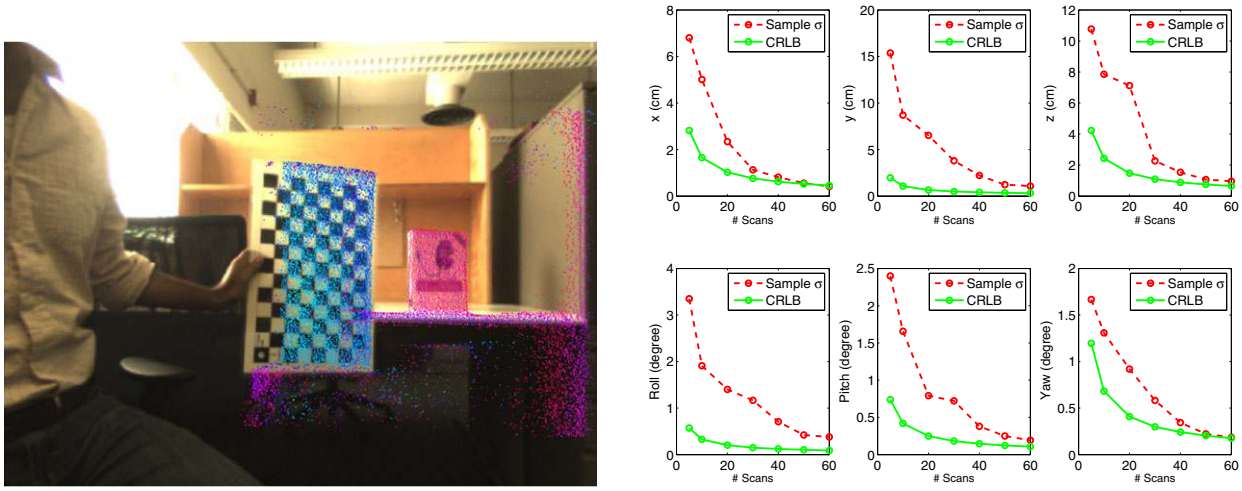
Figure 17. Data obtained from a 3D TOF camera and monocular camera system.

4.2. Time-of-flight 3D Camera and Monocular Camera

In this section, we present results from data collected from a 3D time-of-flight (TOF) camera (Xu et al., 2005) and a monocular camera (Point Grey Research Inc., 2010) mounted on a rigid bar [see Figure 17(a)]. A sample image obtained from the monocular camera is shown in 17(d) and the corresponding depth and intensity map of the scene obtained from the TOF 3D camera are shown in Figures 17(b) and 17(c), respectively. The size of the depth map obtained from the 3D camera is 200×200 pixels, which equates to 40,000 3D points per scan. We use the 3D points with the intensity information along with the camera imagery to estimate the calibration parameters within the proposed framework. We assume that the intrinsic calibration parameters of the monocular camera are either known or are pre-computed using any standard method [e.g., (Zhang, 2000)]. In Figure 18, we show qualitative calibration results for projecting the 3D points onto the corresponding camera imagery using the estimated rigid-body transformation. We also show how the calibration results improve when multiple scans are considered in the MI-based calculation. We observe that the standard deviation of the estimated calibration parameters decreases and approaches the CRLB as the number of scans used to calculate the MI is increased.

4.3. 2D Laser Scanner and Monocular Camera

In this section, we present results from data collected from a 2D laser scanner (Hokuyo, 2009) and a monocular camera (Point Grey Research Inc., 2010) mounted on a rigid bar, as shown in Figure 19. This type of sensor setup is typical for an indoor SLAM problem. In our case, the single-beam 2D laser scanner operates at 30 Hz and provides 540 points per scan, hence the number of scans required to achieve small variance of the calibration parameters is significantly large as compared to the Velodyne (of the order of a few hundred scans). The quality of the minimum variance estimate (calculated from 700 scans) is shown in Figure 19(b). Although we have used up to 700 scans in this experiment, the total number of points used to estimate the MI still remains significantly less ($700 \times 540 = 378,000$ points) as compared to 20 Velodyne scans (i.e., $20 \times 80,000 = 1,600,000$ points). In Figure 19(c), we plot the sample standard deviation of the estimated calibration parameters and the corresponding CRLB as a function of the number of scan pairs used. As observed in the earlier experiments (Sections 4.1.2 and 4.2), we see a decrease in parameter variance as the number of scans is increased and the sample standard deviation approaches the value predicted by the CRLB. Although, the standard deviation does not seem to converge to zero, if we use more scans the standard deviation can be further reduced and will converge to zero in the limiting case.



(a) TOF point cloud projected onto the camera imagery

(b) MI-based calibration result

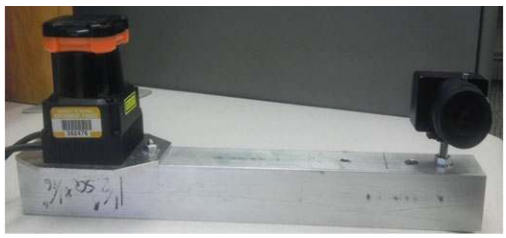
Figure 18. Results for the MI-based calibration of a 3D TOF camera and a monocular camera. (a) TOF point cloud projected onto the camera imagery; the points are color-coded based on scene depth from the camera. (b) We plot the uncertainty of the recovered calibration parameter versus the number of scans used. The red (dashed line) plot shows the sample-based standard deviation (σ) of the estimated calibration parameters calculated over 1,000 trials. The green (solid line) 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.

It should be noted that in this experiment, we have not mounted the sensors to any robotic platform; instead, we have attached the sensors on a rigid bar and moved the whole assembly in space while collecting the data for calibration (i.e., there is no restriction of movement). However, if the sensors were to be mounted on a planar robot, observability of certain calibration parameters will be an issue. We think that this issue can be resolved, however, by designing a rigid fixture for the sensors (e.g., a rigid bar, as shown in the experiment) and computing the relative sensor transform using the method described here before mounting the fixture to the robot, for example.

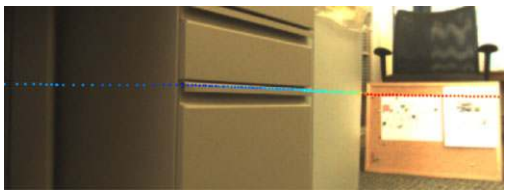
5. DISCUSSION

The MI-based framework for calibration of multimodal sensors presented here assumes that the range sensor also provides reflectivity of the surface apart from the range information. However, oftentimes we need to use other sensing modalities (e.g., sonar or laser without reflectivity) due to system constraints or for certain specific requirements. We have not tested the proposed MI-based framework with any sensors that do not provide a direct correlation as observed between reflectivity and gray-scale values. However, we believe that 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 (i.e., 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 (Figure 20). The MI between these features should exhibit a maxima at the sought-after rigid-body transformation. There might be other ways to extract highly correlative features for such sensors, and it will be worthwhile to explore the use of these features in the MI-based calibration framework.

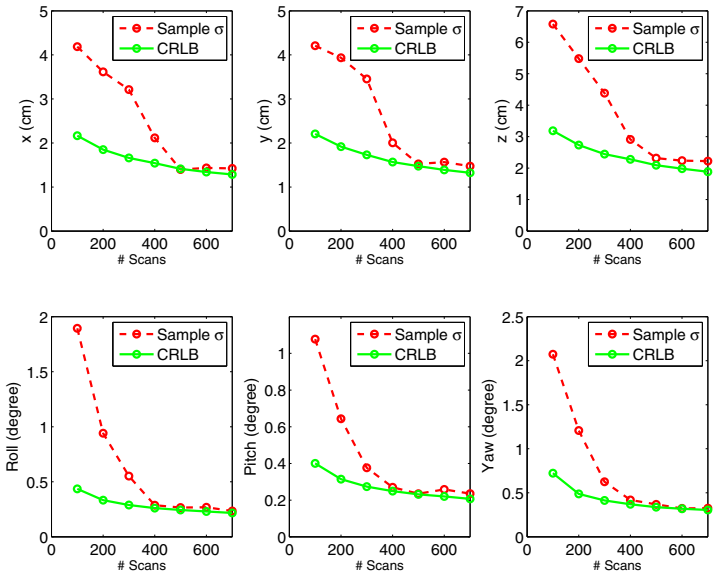
Additionally, the proposed algorithm is formulated as an optimization problem that maximizes a cost function to estimate the unknown calibration parameters. Therefore, the optimization techniques used to solve the unknown variables directly affect the robustness and computational complexity of the algorithms. In this paper, we have used a simple gradient descent technique to solve the optimization problem. It works well when we use sufficient data, however it is sensitive to initialization errors and has the tendency to get trapped in a local optima. Moreover, since the cost function is a nonparametric function of the unknown variables, the gradient is computed numerically, thereby making it computationally expensive and inaccurate. Several methods of directly estimating the derivative of the MI-based cost function using interpolation techniques (Maes et al., 1999; Panin & Knoll, 2008) have been developed in the past. In the future, we would like to use these techniques to obtain a better estimate of the first- and second-order derivatives of the MI-based cost function, thereby improving the estimated calibration parameters and the corresponding CRLB.



(a) 2D laser/camera setup



(b) 2D lidar points projected onto camera image



(c) Multi-view calibration result

Figure 19. Results for the MI-based calibration of a 2D lidar and a monocular camera. (a) 2D laser scanner and a monocular camera mounted on a horizontal bar. (b) 2D lidar points projected onto the camera image using the estimated transform. Points are color-coded based on distance from the camera: blue–close, red–far. (c) We plot the uncertainty of the recovered calibration parameter versus the number of scans used. The red (dashed line) plot shows the sample-based standard deviation (σ) of the estimated calibration parameters calculated over 1,000 trials. The green (solid line) 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.

6. CONCLUSIONS

We presented an information theoretic algorithm to automatically estimate the rigid-body transformation between a camera and a lidar range sensor. It is important to note that the reflectivity of the 3D points obtained from the range sensor and the intensity of the pixel obtained from the camera are discrete signals generated by sampling the same physical scene, but in a different manner. Since the underlying structure generating these signals is common, they are statistically dependent upon each other. We use MI as the measure of this statistical dependence and formulate a cost function that is maximized for the correct calibration parameters. The source code of an implementation of the proposed algorithm in C++ is available for download from our server at http://robots.engin.umich.edu/SoftwareData/ExtrinsicCalib.

The proposed algorithm is completely data-driven and does not require any artificial targets to be placed in the field-of-view of the sensors, making it fairly easy to calibrate. Target-based methods, on the other hand, require special fiducials to be placed in the environment, which is onerous. This is the reason why sensor calibration in a robotic application is typically 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, the assumption 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 setting up 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.

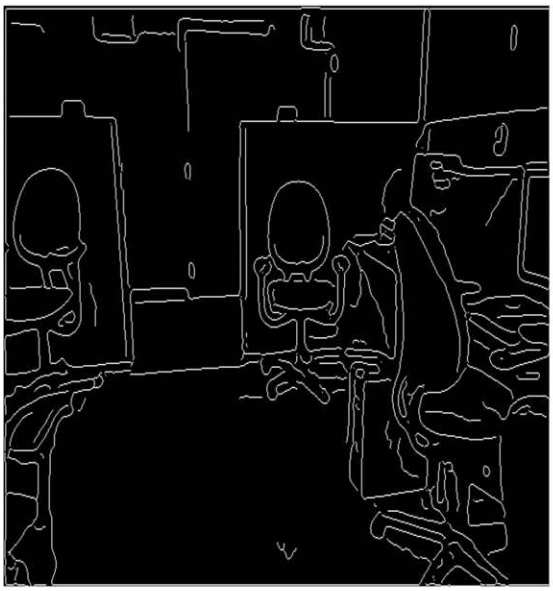
We compared the proposed algorithm with two targetless calibration methods. The first method, from Williams et al. (2004), is fundamentally very similar to the proposed method as it uses correlation of lidar reflectivity and grayscale intensity values from the camera to estimate the calibration parameters. Since χ^2 cost and MI are both statistical measures of the correlation of these intensity values, we observe similar results when the same number of scans is used to estimate the calibration parameters from the χ^2-test and the proposed method. The advantage of using MI is that it is a well-studied measure, and several methods of robust and efficient estimation of MI from the sample data have already been developed, which can be directly used within the proposed framework. The second method, that of Levinson and Thrun (2012), also uses the joint statistics of data

(a) Monocular Camera

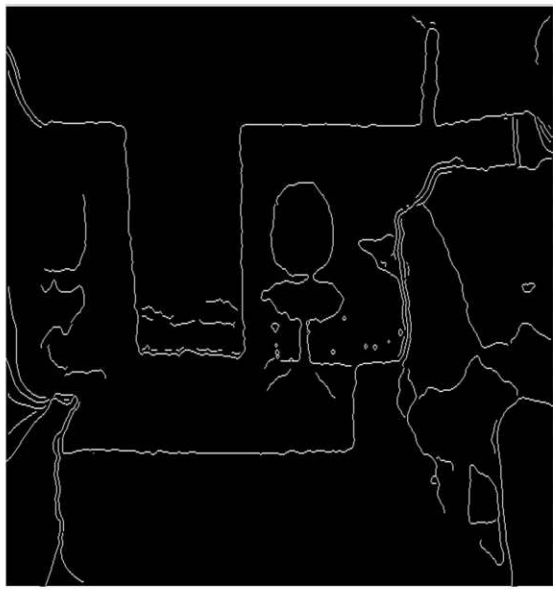
(b) Kinect Camera

(c) Color image

(d) Depth map



(e) Edges from color image



(f) Edges from depth map

Figure 20. Here we illustrate an example extension of the MI-based calibration framework to a monocular camera with a Kinect camera, which does not provide any reflectivity information. The color image and the corresponding depth map from the Kinect camera are shown below (center panel). The edges extracted from the color image and the corresponding depth map (bottom panel) clearly show a correlation.

obtained from the lidar and camera to estimate the calibration parameters. Although we observe similar trends in results, the proposed method produces better results with smaller amounts of data. This is mainly because Levinson's method discards a significant amount of data in the preprocessing stage. An important advantage of Levinson's method, however, is that it does not use reflectivity values from lidar data and, therefore, can be used to calibrate sensors that do not necessarily provide reflectivity information.

We showed that the proposed algorithm works with a wide variety of sensors commonly used in indoor/outdoor robotics. Various experiments were performed to show the robustness and accuracy of the algorithm in typical robotics applications. Whether it is a 3D laser scanner and an omnidirectional camera system mounted on the roof of a car, or a 2D laser scanner and a monocular camera mounted on a robotic platform for indoor applications, the proposed method works equally well.

Our algorithm also provides a measure of the uncertainty of the estimated parameters through the Cramér-Rao lower bound. We have shown in our experiments that the sample variance of the estimated parameters approaches the CRLB as the number of scans is increased, therefore in the limit our estimator can be considered to be an *efficient* estimator. Moreover, in the limiting case, the CRLB can be considered as the true variance of the estimated parameters and can be readily used within any probabilistic robotics perception framework.

ACKNOWLEDGMENTS

This work was supported by Ford Motor Company via a grant from the Ford-UofM alliance via Award No. N015392.

REFERENCES

Alempijevic, A., Kodagoda, S., Underwood, J. P., Kumar, S., & Dissanayake, G. (2006). Mutual information based sensor registration and calibration. In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 25–30), Orlando, FL.

Bao, S. Y., & Savarese, S. (2011). Semantic structure from motion. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2025–2032), Providence, RI.

Barzilai, J., & Borwein, J. M. (1988). Two-point step size gradient methods. IMA Journal of Numerical Analysis, 8, 141–148.

Boughorbal, F., Page, D. L., Dumont, C., & Abidi, M. A. (2000). Registration and integration of multisensor data for photorealistic scene reconstruction. In Proceedings of 28th AIPR Workshop on 3D Visualization for Data Exploration and Decision Making (vol. 3905, pp. 74–84).

Carlevaris-Bianco, N., Mohan, A., McBride, J. R., & Eustice, R. M. (2011). Visual localization in fused image and laser range data. In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 4378–4385), San Francisco.

Chao, A., & Shen, T. J. (2003). Nonparametric estimation of Shannons index of diversity when there are unseen species in sample. Environmental and Ecological Statistics, 10(4), 429–443.

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

Forrest, S. (1993). Genetic algorithms: Principles of natural selection applied to computation. Science, 261(5123), 872–878.

Gong, X., Lin, Y., & Liu, J. (2013). 3D lidar-camera extrinsic calibration using an arbitrary trihedron. Sensors, 13(2), 1902–1918.

Hartley, R., & Zisserman, A. (2000). Multiple view geometry in computer vision. Cambridge University Press.

Hausser, J., & Strimmer, K. (2009). Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. Journal of Machine Learning Research, 10, 1469–1484.

Hill, D., Hawkes, D., Harrison, N., & Ruff, C. (1993). A strategy for automated multimodality image registration incorporating anatomical knowledge and imager characteristics. Information Processing in Medical Imaging, 687, 182–196.

Hokuyo (2009). Scanning range finder: UTM-30LX. Technical report, Hokuyo. Specification sheet and documentations available at www.hokuyo-aut.jp/02sensor/07scanner/utm_30lx.html.

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. Science, 220(4598), 671–680.

Lehmann, E. L., & Casella, G. (2011). Theory of point estimation. Springer Texts in Statistics Series. Springer.

Levenberg, K. (1944). A method for the solution of certain problems in least squares. The Quarterly of Applied Mathematics 2, 164–168.

Levinson, J., & Thrun, S. (2010). Unsupervised calibration for multi-beam lasers. In Proceedings of International Symposium on Experimental Robotics, Delhi, India.

Levinson, J., & Thrun, S. (2012). Automatic calibration of cameras and lasers in arbitrary scenes. In Proceedings of International Symposium on Experimental Robotics, Quebec City, Canada.

Li, G., Liu, Y., Dong, L., Cai, X., & Zhou, D. (2007). An algorithm for extrinsic parameters calibration of a camera and a laser range finder using line features. In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 3854–3859), San Diego, CA.

Li, Y., Ruichek, Y., & Cappelle, C. (2013). Optimal extrinsic calibration between a stereoscopic system and a lidar. IEEE Transactions on Instrumentation and Measurement, 62(8), 2258–2269.

Lin, Y., & Medioni, G. (2008). Mutual information computation and maximization using GPU. In Computer Vision and Pattern Recognition Workshop (pp. 1–6), Anchorage, AK.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., & Suetens, P. (1997). Multimodality image registration by

maximization of mutual information. IEEE Transactions on Medical Imaging, 16, 187–198.

Maes, F., Vandermeulen, D., & Suetens, P. (1999). Comparative evaluation of multiresolution optimisation strategies for multi modality image registration by maximisation of mutual information. Medical Image Analysis, 3(4), 373–386.

Marquadrt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. SIAM Journal on Applied Mathematics 11, 431–441.

McDonald, J. H. (2009). Handbook of biological statistics, 2nd ed. Baltimore, MD: Sparky House Publishing.

Mei, C., & Rives, P. (2006). Calibration between a central catadioptric camera and a laser range finder for robotic applications. In Proceedings of IEEE International Conference on Robotics and Automation (pp. 532–537), Orlando, FL.

Mirzaei, F. M., Kottas, D. G., & Roumeliotis, S. I. (2012). 3D lidar-camera intrinsic and extrinsic calibration: Observability analysis and analytical least squares-based initialization. International Journal of Robotics Research, 31(4), 452–467.

Moghadam, P., Bosse, M., & Zlot, R. (2013). Line-based extrinsic calibration of range and image sensors. In Proceedings of IEEE International Conference on Robotics and Automation (vol. 2, pp. 4–11), Karlsruhe, Germany.

Napier, A., Corke, P., & Newman, P. (2013). Cross-calibration of push-broom 2D lidars and cameras in natural scenes. In Proceedings of IEEE International Conference on Robotics and Automation (pp. 3664–3669), Karlsruhe, Germany.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. The Computer Journal, 7(4), 308–313.

Newman, P., Cole, D., & Ho, K. (2006). Outdoor SLAM using visual appearance and laser ranging. In Proceedings of IEEE International Conference on Robotics and Automation (pp. 1180–1187), Orlando, FL.

Nunnez, P., Jr, P. D., Rocha, R., & Dias, J. (2009). Data fusion calibration for a 3D laser range finder and a camera using inertial data. In Proceedings of the 4th European Conference on Mobile Robots (pp. 31–36), Mlini/Dubrovnik, Croatia.

Pandey, G., McBride, J., Savarese, S., & Eustice, R. M. (2011a). Visually bootstrapped generalized ICP. In Proceedings of IEEE International Conference on Robotics and Automation (pp. 2660–2667), Shanghai, China.

Pandey, G., McBride, J. R., & Eustice, R. M. (2011b). Ford campus vision and lidar data set. International Journal of Robotics Research, 30(13), 1543–1552.

Pandey, G., McBride, J. R., Savarese, S., & Eustice, R. M. (2010). Extrinsic calibration of a 3d laser scanner and an omnidirectional camera. In IFAC Symposium on Intelligent Autonomous Vehicles (vol. 7, pp. 336–341), Leece, Italy.

Pandey, G., McBride, J. R., Savarese, S., & Eustice, R. M. (2012a). Automatic targetless extrinsic calibration of a 3D lidar and camera by maximizing mutual information. In Proceedings of AAAI National Conference of Artificial Intelligence (pp. 2053–2059), Toronto, Canada.

Pandey, G., McBride, J. R., Savarese, S., & Eustice, R. M. (2012b). Toward mutual information based automatic registration of 3D point clouds. In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 2698–2704), Algarve, Portugal.

Panin, G., & Knoll, A. (2008). Mutual information-based 3D object tracking. International Journal of Computer Vision, 78(1), 107–118.

Point Grey Research Inc. (2009). Spherical vision products: Ladybug3. Technical report, Point Grey Research. Specification sheet and documentations available at www.ptgrey.com/products/ladybug3/index.asp.

Point Grey Research Inc. (2010). Imaging products: Firefly IEEE 1394a. Technical report, Point Grey Research. Specification sheet and documentations available at www.ptgrey.com/products/fireflymv/fireflymv_usb_firewire_cmos_camera.asp.

Rodriguez, F., Fremont, V., & Bonnifait, P., et al. (2008). Extrinsic calibration between a multi-layer lidar and a camera. In IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (pp. 214–219).

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

Scott, D. W. (1992). Multivariate density estimation: Theory, practice, and visualization. New York: John Wiley.

Silverman, B. W. (1986). Density estimation for statistics and data analysis. Monographs on Statistics and Applied Probability.

Smith, R., Self, M., & Cheeseman, P. (1988). A stochastic map for uncertain spatial relationships. In Proceedings of International Symposium on Robotics Research (pp. 467–474), Santa Clara, CA.

Tamjidi, A., & Ye, C. (2012). 6-DOF pose estimation of an autonomous car by visual feature correspondence and tracking. International Journal of Intelligent Control and Systems, 17(3), 94–101.

Taylor, Z., & Nieto, J. (2012). A mutual information approach to automatic calibration of camera and lidar in antural environments. In Proceedings of Australian Conference on Robotics and Automation, Wellington, New Zealand.

Unnikrishnan, R., & 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. Technical report, Velodyne. Available at www.velodyne.com/lidar/products/white_paper.

Viola, P., & Wells, W. (1997). Alignment by maximization of mutual information. International Journal of Computer Vision, 24, 137–154.

Wang, R., Ferrie, F., & Macfarlane, J. (2012). Automatic registration of mobile lidar and spherical panoramas. In Computer Vision and Pattern Recognition Workshops (CVPRW),

IEEE Computer Society Conference (pp. 33–40), Providence, RI.

Wenzel, W., & Hamacher, K. (1999). A stochastic tunneling approach for global minimization. Physical Review Letters, 82(15), 3003–3007.

Whittaker, E. T., & Robinson, G. (1967). The Newton-Raphson method. The calculus of observations: A Treatise on Numerical Mathematics, 44(4), 84–87.

Williams, N., Low, K. L., Hantak, C., Pollefeys, M., & Lastra, A. (2004). Automatic image alignment for 3d environment modeling. In Proceedings of IEEE Brazilian Symposium on Computer Graphics and Image Processing (pp. 388–395).

Woods, R. P., Mazziotta, J. C., & Cherry, S. R. (1993). MRI-PET registration with automated algorithm. Journal of Computer Assisted Tomography, 17(4), 536–546.

Xu, Z., Schwarte, R., Heinol, H., Buxbaum, B., & Ringbeck, T. (2005). Smart pixel—Photonic mixer device (PMD) new system concept of a 3D-imaging camera-on-a-chip. Technical report, PMD Technologies.

Zhang, Q. (2004). Extrinsic calibration of a camera and laser range finder. In Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (pp. 2301–2306), Sendai, Japan.

Zhang, Z. (2000). A flexible new technique for camera calibration. IEEE Transaction on Pattern Analysis and Machine Intelligence, 22(11), 1330–1334.

Zhou, L., & Deng, Z. (2012). Extrinsic calibration of a camera and a lidar based on decoupling the rotation from the translation. In Proceedings of IEEE Intelligent Vehicles Symposium (IV) (pp. 642–648).