Optics and Lasers in Engineering 51 (2013) 394–401

Contents lists available at SciVerse ScienceDirect

Optics and Lasers in Engineering

journal homepage: www.elsevier.com/locate/optlaseng

Extrinsic calibration of a 3D LIDAR and a camera using a trihedron

Xiaojin Gong*, Ying Lin, Jilin Liu

Department of Information Science and Electronic Engineering, Zhejiang University, Hangzhou, Zhejiang 310027, PR China

ARTICLE INFO

Article history:
Received 14 September 2012
Received in revised form
23 October 2012
Accepted 30 November 2012
Available online 20 December 2012

Keywords:
Extrinsic calibration
3D LIDAR
Camera
Sensor fusion

ABSTRACT

In this paper, we propose a novel method to easily conduct the extrinsic calibration between a camera and a 3D LIDAR. By taking advantage of orthogonal trihedrons which are ubiquitous in structured environments, our method makes it convenient for a mobile robot to collect the data needed for calibration. The proposed method estimates the relative position of the sensors by first determining the pose of each plane of the trihedron in two acquisitions and then solving a planarity-constrained optimization problem. The algorithm is validated first by performing experiments on simulated data and investigating its sensitivity to noise. Moreover, we carry out a calibration for a real 3D LIDAR–camera system and further apply the calibration results to render 3D points with pixels' colors. The colored 3D points visually validate the effectiveness of our method.

1. Introduction

Nowadays, 3D light detection and ranging sensors (LIDARs) are commonly used in conjunction with cameras on mobile robots for simultaneously localization and mapping (SLAM) [10], obstacle detection [4], object recognition [7], and other navigation tasks. These applications prefer to take advantage of both range and visual information simultaneously. Thus, sensor fusion is needed. It first requires us to estimate the relative pose between two sensors, namely extrinsic calibration. This paper aims to address this fundamental problem in a convenient way.

In this paper, we propose a novel way to extrinsically calibrate a system consisting of a 3D LIDAR and a camera. The proposed method distinguishes itself from most published techniques in three aspects:

1. It chooses a trihedron that is composed of three orthogonal planes as a calibration rig. Such kind calibration setting is ubiquitous in both indoor and outdoor environments. In addition, compared to other calibration rigs such as planar checkerboards [19,17], this setting is less easy to be perturbed or altered even under heavy wind or other severe weather conditions.
2. During calibration, it only requires a user to mark out four specific points in images and select the region of each plane in both images and 3D point clouds. Hence, only a couple of points are needed and the precision of points for region selection does not matter much.
3. It only needs two LIDAR–camera acquisitions of the trihedron. Based upon the two acquisitions, the proposed method is able to achieve good calibration results.

2. Related work

A camera provides us with dense color images and a laser range finder produces sparse, but accurate, range information. These two sensors offer different advantages so that sensor fusion is desirable in many applications. The quality of fusion highly depends upon the assessment of the sensors' extrinsic parameters. Early work on extrinsic calibration mainly focuses on 2D LIDARs and perspective cameras [19]. Along with the advancement of 3D range sensing techniques, particularly the development of Velodyne HDL-64E LIDARs [18], various calibration methods [13,14] for 3D LIDAR and camera systems are published in recent years.

Roughly speaking, extrinsic calibration techniques for LIDAR–camera systems can be categorized in terms of the geometric constraints upon which they rely. For instance, analogous to camera intrinsic calibration [20], Mei and Rives [9] and Scaramuzza et al. [16] employ point correspondences obtained in both images and 3D point clouds to estimate the rigid transformation between two sensors. Although they propose automatic means for interest point detection in range data, it is hard to locate points in range as precisely as that in visual images due to the lower and non-uniform resolutions. Line [2,11] and circle [15] correspondences are therefore considered in later methods to

* Corresponding author. Tel./fax: +86 571 87951529.
E-mail addresses: gongxj@zju.edu.cn (X. Gong), ly86@zju.edu.cn (Y. Lin), liujl@zju.edu.cn (J. Liu).

0143-8166/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.optlaseng.2012.11.015

circumvent the uncertainty in point localization. Luiz et al. [10,1] and Nunez et al. [12] also make use of inertial sensors to reduce the number of point correspondences needed.

A more common group of the techniques [19,13,17] employ a planar rig for calibration. The rig is covered with a checkerboard pattern with a known size which provides a metric reference. Imaged corners on the checkerboard pattern are used to estimate the poses of the plane in camera frames. Meanwhile, 3D points falling on the plane are taken into account to estimate the poses in the LIDAR coordinates. With a couple of LIDAR–camera acquisitions of the calibration rig, it is able to assess the relative transformation between the two sensors. Zhang and Pless [19] first implement this idea to calibrate a 2D range finder and a camera. Unnikrishnan [17] and Pandey et al. [13] further extend this method to systems composed of a 3D LIDAR and a camera. Unnikrishnan [17] also makes a toolbox available online.

In addition to the above-mentioned approaches, it is worth mentioning a method [14] that is proposed very recently. Different from all others, the method incorporates the laser reflectivity of 3D LIDAR points for calibration. By maximizing the mutual information of the reflectivity and the intensity of corresponding points in both sensors, it estimates the relative transformation without any calibration rigs. Instead of detecting points in range data, the point correspondences are built by projecting 3D LIDAR points onto images so that the previous mentioned localization noise does not bother. This method is novel and works well when the laser reflectivity is available, although it takes minutes for the iterative optimization to get converged.

In this paper, we do not assume the availability of the reflectivity information from a LIDAR. Like all except [14], our method takes into account only visual images and range data, and focuses on how to calibrate two sensors more conveniently. Hence, the proposed method can be applied to an arbitrary range finder no matter if it produces reflectivity information or not.

3. Methodology

Our method relies upon the geometric constraints of an orthogonal trihedron in two LIDAR–camera acquisitions. The trihedron consists of three planes which are orthogonal to each other. Such kind of object is ubiquitous in natural scenes, for instance, two adjacent walls of a building together with the ground plane, in either indoor or outdoor environments. Hence it is handy for a mobile robot to capture data for calibration.

In this section, we present the formulation of the extrinsic calibration problem with respect to a trihedral calibration rig in two views and introduce the calibration procedure in detail.

3.1. Problem formulation

Let us consider a pre-calibrated camera, meaning that its intrinsic parameters are known, and a 3D LIDAR rigidly assembled on a robot. A trihedron is observed simultaneously by two sensors. Assuming that the trihedron is of three strictly orthogonal planes, we set up the world coordinate system in such a way as illustrated in Fig. 1(c). The origin is the intersection point of the planes, Z-axis is aligned with the intersection line of two upright planes, X- and Y-axes are aligned with the other two intersection lines so that the right-hand rule is followed. The LIDAR coordinate system is defined so that its origin is the convergence point of the laser beams and the axes are aligned with the laser beams as shown in Fig. 1(b). Moreover, the camera coordinate system is set up to be coincident with the one defined on its own projection model [6]. The rotation and translation from the world coordinate system to the camera frame and the LIDAR system are, respectively, denoted as $\mathbf{R}_c,\mathbf{T}_c$, and $\mathbf{R}_l,\mathbf{T}_l$. We also denote the rotation and translation from the LIDAR coordinate system to the camera as $\mathbf{R},\mathbf{T}$, which are the parameters we aim to estimate in the calibration task. Here, $\mathbf{R}_c$, $\mathbf{R}_l$, $\mathbf{R} \in SO(3)$, and $\mathbf{T}_c,\mathbf{T}_l,\mathbf{T}$ are 3D column vectors.

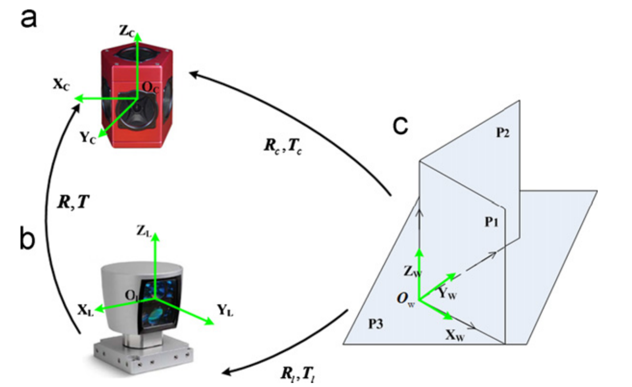

Fig. 1. A calibration model. In this configuration, a Ladybug3 camera (a) and a Velodyne HDL-64E LIDAR (b) are rigidly assembled together. A trihedral calibration rig (c) models an outdoor building scenario. This calibration model fits to indoor trihedral scenarios as well.

A plane in a coordinate system can be represented by $\mathbf{N}^T\mathbf{p}-d=0$, where $\mathbf{p}$ is an arbitrary 3D point lying on the plane, $\mathbf{N}$ indicates the plane's normal vector, and d is the distance to the origin. Before going into the detail, we make clear some notations first. We use $(\mathbf{N}_{s_i}^j, d_{s_i}^j)$ to represent the pose of a specific plane, where $s \in \{c,l\}$ indicates the sensor coordinate system to which the plane is relative, c for camera and l for LIDAR. $i \in \{1,2\}$ indicates the index of a view, and $j \in \{1,2,3\}$ denotes the index of a plane of the trihedron. As shown in Fig. 1(c), the first plane is spanned by the X- and Z-axes, the second one is spanned by the Y- and Z-axes, and the remaining is indexed by 3. In addition, we take $\mathbf{p}_{s_i}^{j,k}$ to indicate the kth point on a specific plane.

Now, let us consider two LIDAR–camera acquisitions. Assuming that the poses of all planes are known, we can formulate the extrinsic calibration task as an optimization problem. In terms of the planarity constraints we get the form

$$\arg\min_{\mathbf{R},\mathbf{T}} \sum_{i=1}^{2}\sum_{j=1}^{3}\sum_{k=1}^{M(i,j)} \|\mathbf{N}_{c_i}^{j}{}^T(\mathbf{R}\mathbf{p}_{l_i}^{j,k}+\mathbf{T})-d_{c_i}^{j}\|^2, \tag{1}$$

where $M(i,j)$ is the number of points on a corresponding plane. This formulation indicates that all points on a plane in the LIDAR coordinates should be coplanar also in the camera coordinate system.

It is a non-linear optimization problem which can be solved by Levenberg–Marquardt method [8]. However, it requires a good initialization of $\mathbf{R}$ and $\mathbf{T}$, and before solving the problem, we need to determine $\mathbf{N}$ and d for all planes. All of these are detailed below.

3.2. Extrinsic LIDAR calibration

Extrinsic LIDAR calibration is to estimate the transformation, $\mathbf{R}_l$ and $\mathbf{T}_l$, from the world coordinate system the LIDAR. As byproducts, the pose of each plane is also determined. For the sake of simplicity, we here drop off the i subscript and the j superscript, and take one plane as an example. When a user selects a region of 3D points that mostly lie on a plane, the plane's pose $(\mathbf{N}_l,d_l)$ is estimated by

$$\arg\min_{\mathbf{N}_l,d_l} \sum_{k=1}^{M} \|\mathbf{N}_l^T\mathbf{p}_l^k-d_l\|^2. \tag{2}$$