

Estimating Extrinsic Parameters Between a Stereo Rig and a Multi-Layer Lidar Using Plane Matching and Circle Feature Extraction

Trevor Gee¹, Jason James¹, Wannes Van Der Mark¹, Alfonso Gastelum Strozzi², Patrice Delmas¹, and Georgy Gimel'farb¹

¹Department of Computer Science, University of Auckland, New Zealand

²CCADET-HGM. Universidad Nacional Autónoma de México. Ciudad de México

Abstract

In this work, we investigate the problem of estimating a rigid transform mapping between a calibrated stereo camera rig and a multi-layer lidar. Such a transform may be used to merge data between these 2 systems, addressing the colourless sparse nature of the lidar data and potentially improving depth estimation from the stereo pairs. The proposed approach features a novel planar calibration object with three circular features allowing for the robust acquisition of corresponding features between sensors. A closed-form registration of correspondences is proposed, leading to the derivation of the required transform. The main appeal of the proposed approach is its conceptually simple formulation and the fact that only a single image from each device is required for calibration. Our experiments were performed on real data captured in outdoor and indoor environments and demonstrate good performance with a Velodyne VLP-16 lidar and GOPRO HERO 3+ Stereo rig.

1 Introduction

With the advent of pervasive computer vision and high performance lightweight energy sources, a wealth of fields have opened to machine vision systems, most notably in forestry [1] and farming [2]. Of great interest are affordable UAV-based and multi-camera band systems as they have the potential to produce more robust and complete data-sets. As such, the fusion of active (say lidar) and passive vision (say stereo-vision) apparatus provides a unique opportunity to generate better quality computer models. Lidar data may gain colour and density, while point clouds from stereo images may gain accuracy and points from a wider field of view. As such, a critical step is finding the mapping between a calibrated stereo rig and a multi-layer lidar. Typically this mapping is a rigid transform between coordinate systems.

The main obstacles in determination of this mapping are (i) the sparse nature of lidar data, which leads to difficulties in finding exact locations of corresponding features and measurement noise; (ii) the noise associated with a stereo system rig; (iii) the distance from the scene; and (iv) the stereo-matching algorithm performance which depends on the image content.

The main advantages are (i) the very dense nature of depth-generated from stereo-vision systems associated with a reasonably high accuracy (as below lidar accuracy) at close range; (ii) the illumination invariant

nature of lidar data sampling; (iii) the low measurement noise in the Velodyne VLP-16 lidar (used in our experiments) i.e. RMSE 30 mm [3] for a distance up to 100 meters.

Our approach firstly determines a set corresponding features robustly, and then calculates a rigid transform to register those features. The novelty in our approach is a focus on acquiring robust corresponding features, which then greatly simplifies the calculations with respect to feature registration, since much of the uncertainty has been mitigated. This is in contrast with most state-of-the-art algorithms that rely on less robust feature extraction, but then need to optimize solutions in complex error spaces [4] or sample over multiple images [5].

In section 2 we will discuss some of the history with respect to the evolution of the lidar and look at the current state-of-the-art approaches to calibrating them. In section 3 we introduce our experimental rig and calibration object, describe the proposed algorithm. In section 4 we outline and present the results of our experiments and conclude with section 5.

2 Literature Review

Initially, the first lidars typically featured a single 360 scan-line. They were typically used in robotics to automatically assess distance from surrounding obstacles. Early work has been done to attempt to calibrate these types of cameras with colour images, such as the multiple plane approach featured in [5]. However lidars have evolved a lot and now feature multiple scan-lines and are used in many sophisticated projects including the Google car [6].

Calibration of stereo rigs and lidars is a fairly popular area of research and therefore there exists several different approaches that authors have used [7, 8, 9, 4, 10, 11].

Once such approach makes use of external sensors such as the Inertial Measurement Unit (IMU) in conjunction with a planar calibration object in order to establish a set of images and corresponding transformations to constraint the determination of a rigid transform between lidar space and camera space [7] [8].

Another approach uses typical calibration objects used in camera calibration to calibrate lidars. Most of these approaches make use of the calibration object used in Zhang calibration can involve the acquisition of multiple images. The calibration objects are typically detected as planes in lidar space.

This work falls under lidar calibration with specialized calibration objects' category. In [9], a black line on a white sheet of paper is used. In [4] planar calibration object with holes of various geometric shapes are used, which is similar to our calibration object. In [10] a single planar object with a single large circular hole is used. In [11], a set of boxes is used with different coloured surfaces, arranged so that each is touching.

3 Methodology

3.1 Equipment

Our camera rig consists of a Velodyne VLP-16 lidar and 2 GOPRO HERO 3+ Black edition cameras as depicted in figure 1. The GOPRO cameras are synchronized with each lidar via a GOPRO synchronization cable designed for the HERO 3+ and is synchronized with the lidar via software.



Figure 1. (Left) Our experimental acquisition system feature a Velodyne Lidar VLP-16 and two Synchronized GOPRO HERO 3+ Black edition cameras. (Right) Our planar calibration object.

The Velodyne VLP-16 lidar consist of 16 scan-lines, has a 100 meter range and a 360 horizontal field of view and a 30 vertical field of view. Points are extracted with the coordinate system orientated so that the Y-axis is forwards, the X-axis is to the side and the lasers revolve around the Z-axis.

The GOPRO camera system was set to capture single frames of size 4000×3000 pixels with a horizontal field of view of 122.6 degrees and vertical field of view of 94.4 degrees. The baseline between the cameras is 100 mm.

The calibration object chosen (depicted in figure 1) consists of a cardboard plane maintained rigidly with a custom-made aluminium rectangular frame. Its size was carefully chosen to contain the 16 scan-lines of the lidar at a distance of 1 meter, thus making the dimensions 600×850 mm. The white colour makes the calibration object easy to segment via thresholding when placed against a dark background. The choice of three large holes was a compromise between having enough feature point measurements and having enough scan-lines intersecting the feature for identification. The radius of the circle features was chosen to be 150 mm.

3.2 Feature Detection

The extracted features of the calibration object (figure 1), that are used by the alignment algorithm are, firstly the plane, P , and the set of circular feature centers, $C = \{c_1, c_2, c_3\}$ where c_i is a center of a hole within the calibration object for $i = \{1, 2, 3\}$. Here P may be represented as $n \cdot p = d$ where n is the 3-D normal vector to the plane, p is a point in 3-D space and d is a scalar interpreted as an offset to the origin.

3.2.1 Stereo System Calibration Plane Feature detection

The process of extracting features of calibration object from the stereo images is as follows:

1. A depth map is generated from the stereo pair of images. This starts with determining the cameras' calibration parameters and continues with distortion removal, image rectification, stereo matching and a disparity-to-depth conversion as outlined in [12].
2. An image segmentation process (facilitated by the white plane placed against a dark background and adaptive thresholding) extracts a mask of the pixels belonging to the calibration plane.
3. The plane P_{stereo} is then determined by applying the acquired mask to the depth map, computing the 3-D locations of the pixels making up the calibration object and applying a basic RANSAC [13] plane fitting algorithm.
4. The coordinates of the plane are remapped back to 2-D. This is done by defining an arbitrary orthogonal basis within the plane (any vector between 2 unique points on the plane and the cross product of that vector with the normal of the plane). Once a coordinate system is defined, a mapping function can be formed using the scalar product. The main motivation for creating this mapping is to eliminate the foreshortening of the circles that occurs when the camera is not orthogonal to the calibration object (this process is equivalent to rotating the calibration object's orientation so that it is orthogonal to the image plane). This foreshortening may otherwise effect the quality of our subsequent circle detection.
5. Given the new 2-D mapping of the calibration object image, feature center positions, C_{stereo} , are found using a Canny edge detector followed by a Hough transform [14] and are further refined using an approach similar to the Levenberg-Marquardt in alternative space as mentioned in [15].

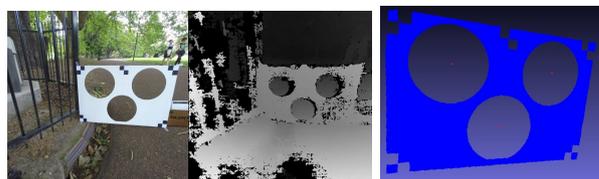


Figure 2. (Left) A rectified color image with distortion removed. (Center) The corresponding disparity map. (Right) 3D planar calibration object mask with circle centers (in red).

3.2.2 Lidar Calibration Plane Feature Detection

The process of extracting features of the calibration object from the lidar data is as follows:

1. The Lidar points are filtered into a relevant subset by determining which of the above points project onto a forward facing virtual image plane applying the projection transform computed for the stereo-vision system as described in section 3.2.1 and exemplified in figure 3.
2. Knowing that the calibration object is approximately 1 meter from the lidar allows the calibration object plane, P_{lidar} , to be determined via a RANSAC plane fit.
3. Knowing the P_{lidar} and size of the calibration object allows for the determination of its boundaries and the determination of three clusters (using k-means) of pixels representing the holes in the calibration object. The centers, C_{lidar} , may then be approximated as center of gravity of these clusters.
4. To refine C_{lidar} , we transform the point cloud into 2-D space with respect to two arbitrary orthogonal axis in the calibration project plane, representing each point as a binary value (belonging either to background or foreground). A RANSAC-based circle fitting algorithm identifies circle edges before final refinement using the same approach as in step 5 of section in section 3.2.1

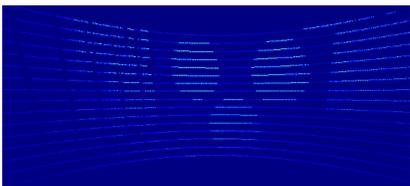


Figure 3. Lidar front points projected onto an image plane and colored with respect to depth.

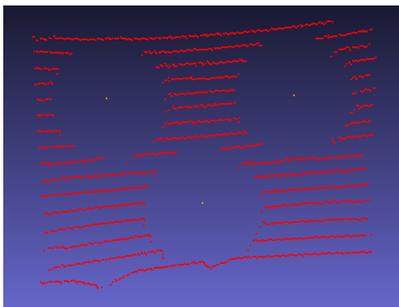


Figure 4. Lidar point cloud with detected circle centers

3.3 Feature Alignment

After feature matching, we have reduced the problem to an alignment of calibration objects from the respective sensor spaces as depicted in figure 5. This can be solved using a variety of approaches. Our two-step approach is as follows:

First, the plane P_{stereo} is aligned with the plane P_{lidar} where P_{stereo} is defined as $n_{stereo} \cdot p = d_{stereo}$ and P_{lidar} is defined as $n_{lidar} \cdot p = d_{lidar}$. First, a translation $-n_{stereo}d_{stereo}$ transforms P_{stereo} to the origin.

Next a rotation about angle $\theta = \text{acos}(n_{stereo} \cdot n_{lidar})$ and axis $n_{stereo} \times n_{lidar}$ transforms P_{stereo} to the same orientation as P_{lidar} . Finally the translation $n_{lidar}d_{lidar}$ transforms the plane to its new position.

Next, the feature sets C_{lidar} and C_{stereo} are aligned. This is achieved by aligning centroids of the feature sets and then by aligning their orientations. To align centroids (which we determine using the balance method), we perform the translation determined by $\text{Centroid}(C_{stereo}) - \text{Centroid}(C_{lidar})$. The orientations are aligned by finding the inverse cosine of the average scalar product between corresponding feature vectors relative to the centroid, and rotating relative to the axis n_{lidar} .

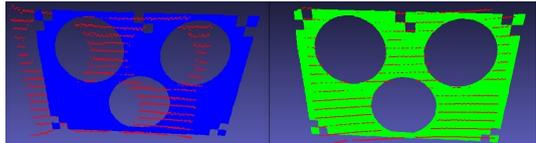


Figure 5. The alignment problem. (Left) unaligned models. (Right) models after alignment.

4 Results

In order to evaluate our work, we introduce two error metrics.

The first of these metrics is the *plane error* defined as the average shortest distance between the measured coordinates in lidar space and the calibration object plane in stereo space. The purpose of this metric is to indicate the quality of the plane alignment with respect to the registration. An average error close to the encountered noise in lidar space (about 30 mm) typically indicates a reasonably close alignment.

The formulation of the second metric attempts to quantify the quality of the alignment of the circular hole features between lidar and GOPRO. We define it as the measure of the number of calibration object points, in lidar space, that project into the holes of the calibration object in stereo camera space. We have chosen to name this metric *recall* (because it is similar in formulation to the recall metric in binary information retrieval) defined as $\text{recall} = tp/(tp + fn)$, where tp is the number of pixels classified as part of the calibration object and fn is the number projected into holes. Clearly a recall of 1 is optimal.

4.1 Experiments

In order to assess our work we conducted five different trials in various outdoor and indoor environments. The goal of the work was to assess the consistency and repeat-ability of our results in a number of environments. In order to compare the performance of our algorithm against a state-of-the-art algorithm, we implemented a MATLAB version of the work described in [4] (referred to here as Velas). This algorithm was chosen due to its similarity to our approach in terms of calibration object type and use of Velodyne lidar.

During calibration, the calibration object was placed roughly parallel to camera rig, at distance of approximately 1 meter in front of the GOPRO cameras. The

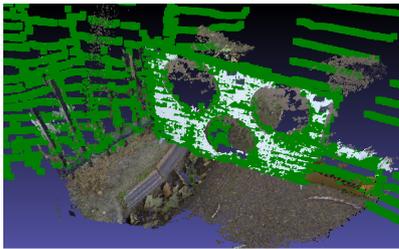


Figure 6. Lidar point cloud (green) and 3-D model (coloured) combined.

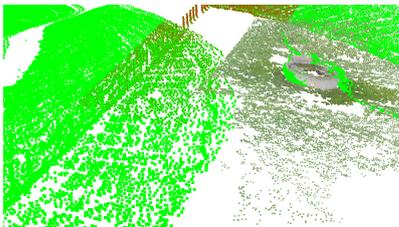


Figure 7. Several lidar point clouds (green) combined with a 3-D model (coloured) of a water trough

parallel orientation is to maximize the detection of the circular holes.

Our findings revealed a general consistency between our results the Velas approach. However a key advantage of our approach is its computational efficiency, as there was no need to compute an inverse distance transform and 6-D search for an optimum transform. The summarized combined results of the 5 experiments are shown in table 1 in terms of the Recall and Plane Error described above.

Table 1. A summary of results of lidar alignment experiments

| Approach | Plane Error | Recall |
|----------|-------------------|--------|
| None | 98.82 mm | 0.76 |
| Velas | 41.8 ± 5.2 mm | 0.96 |
| Proposed | 38.7 ± 6.1 mm | 0.97 |

5 Conclusions

In this work, we set out to verify a robust approach to aligning lidar point clouds with a stereo camera system via a commonly identifiable calibration object. The approach is novel in its simplicity and focus on robust feature extraction over robust pose estimation. Our experimental results demonstrates that our approach is comparable with current state-of-the-art techniques.

We are currently using this calibration strategy for research into registering multiple lidar points clouds against stereo point clouds with good results. An example output from this work may be seen in figure 7.

References

[1] M Karpina, M Jarzabek-Rychard, P Tymkow, and A Borkowski. Uav-based automatic tree growth mea-

surement for biomass estimation. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 685–688, 2016.

- [2] Sebastian Candiago, Fabio Remondino, Marco Dubbini, and Mario Gattelli. Evaluating multispectral images and vegetation indices for precision farming applications from uav images. *Remote Sensing*, 7(4):4026–4047, 2015.
- [3] CL Glennie, A Kusari, and A Facchin. Calibration and stability analysis of the vlp-16 laser scanner. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pages 55–60, 2016.
- [4] Martin Velas, Michal Spanel, Zdenek Materna, and Adam Herout. Calibration of rgb camera with velodyne lidar. 2014.
- [5] You Li, Yassine Ruichek, and Cindy Cappelle. Optimal extrinsic calibration between a stereoscopic system and a lidar. *Transactions on Instrumentation and Measurement*, 62(8):2258–2269, 2013.
- [6] Michelle Birdsall. Google and ite: the road ahead for self-driving cars. *Institute of Transportation Engineers. ITE Journal*, 84(5):36, 2014.
- [7] Hadi Aliakbarpour, Pedro Nunez, Jose Prado, Kamrad Khoshhal, and Jorge Dias. An efficient algorithm for extrinsic calibration between a 3d laser range finder and a stereo camera for surveillance. In *International Conference on Advanced Robotics*, pages 1–6. IEEE, 2009.
- [8] Pedro Núñez, Paulo Drews Jr, Rui P Rocha, and Jorge Dias. Data fusion calibration for a 3d laser range finder and a camera using inertial data. In *ECMR*, pages 31–36, 2009.
- [9] Oleg Naroditsky, Alexander Patterson, and Kostas Daniilidis. Automatic alignment of a camera with a line scan lidar system. In *Robotics and Automation (ICRA), International Conference on*, pages 3429–3434. IEEE, 2011.
- [10] Vincent Fremont, Philippe Bonnifait, et al. Extrinsic calibration between a multi-layer lidar and a camera. In *Multisensor Fusion and Integration for Intelligent Systems. IEEE International Conference on*, pages 214–219. IEEE, 2008.
- [11] Andrew R Willis, Malcolm J Zapata, and James M Conrad. A linear method for calibrating lidar-and-camera systems. In *International Symposium on Modeling, Analysis & Simulation of Computer and Telecommunication Systems*, pages 1–3. IEEE, 2009.
- [12] Trevor Gee, Patrice Delmas, Nick Stones-Havas, Wannes Van Der Mark, Wei Li, Heide Friedrich, and Georgy Gimel’farb. Tsai camera calibration enhanced. In *Machine Vision Applications (MVA), 2015 14th IAPR International Conference on*, pages 435–438. IEEE, 2015.
- [13] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981.
- [14] Priyanka Mukhopadhyay and Bidyut B. Chaudhuri. A survey of hough transform. *Pattern Recognition*, 48(3):993–1010, 2015.
- [15] Nikolai Chernov and Claire Lesort. Least squares fitting of circles. *Journal of Mathematical Imaging and Vision*, 23(3):239–252, 2005.