Parameterless Automatic Extrinsic Calibration of Vehicle Mounted Lidar-Camera Systems

Zachary Taylor and Juan Nieto
University of Sydney, Australia
{z.taylor, j.nieto}@acfr.usyd.edu.au

***Abstract*— This paper presents a new method for automated extrinsic calibration of multi-modal sensors. In particular the paper presents and evaluates a pipeline for calibration of 3D lidar and cameras mounted on a sensor vehicle. Previous methods for multi-modal sensor calibration find the optimal parameters by aligning a set of observations from the different sensor modalities. The main drawback of these methods is the need for a good initialisation in order to avoid converging into a local minima. Our approach eliminates this limitation by combining external observations with motion estimates obtained with the individual sensors. The method operates by utilizing structure from motion based hand-eye calibration to constrain the search space of the optimisation.**

I. Introduction

Sensing redundancy is a pre-requisite for robot robust operation. One of the most difficult problems in building multi-modal representations of the environment is accurate data registration. To integrate the information provided by two sensors, their relative location and orientation must be known. On most mobile robots the sensors are calibrated by either hand labelling points or placing markers such as corner reflectors or chequerboards in the scene. The calibration produced by these methods, while initially accurate, is quickly degraded due to the motion of the robot. For mobile robots working on topologically variable environments across a cascade of spatial scales, as found in agriculture or mining, this can result in significantly degraded calibration after as little as a few hours of operation. For long-term autonomy, calibration parameters have to be determined efficiently and dynamically updated over the lifetime of the robot. This paper presents a pipeline for automatic sensors' extrinsic calibration. Our research is motivated by two broad questions: (i) do we have the right introspection techniques that can provide the robot with tools for self-diagnostic?, (ii) can current sensing approaches provide long-term operation that does not require expert assistance?.

Several markerless automatic calibration methods have been recently proposed. One of the first of these approaches was developed by Levinson et al [1], whose method operates on the principle that depth discontinuities detected by the lidar will tend to lie on edges in the image. Pandey et al. [2] and our previous work [3] presented a method that maximizes the mutual information metric between the lidar's intensity of return and the intensity of the corresponding points in the camera's image. Napier et al. successfully registered a push broom 2D lidar with a camera [4] by first combining the lidar with a navigation solution and then aligning the maginitude of the gradients between the lidar and image. In our own previous work [5] we developed a calibration method that is based on the alignment of the orientation of gradients formed from the lidar and camera. These methods while accurate have a serious limitation in their practicality; they all require an accurate initial guess to the calibration to ensure convergence to a reliable solution.

This initial guess means that some manual measure of the sensors' positions has to be performed every time the sensors' physical configuration is modified. We envision a system where this limitation is removed and a single method can be used to automatically re-calibrate, even when the physical positions of the sensors on the mobile vehicle has changed. The new method we propose is based on the use of the estimated motion of the sensors in combination with hand-eye calibration techniques commonly used in robotic arms to generate an approximate extrinsic calibration. This approximate solution is then used as the initial guess for the more accurate calibration techniques outlined above.

Hand-eye calibration was developed to calibrate the offset between the hand and camera of a robotic arm. In hand-eye calibration, a calibration aid such as a chessboard or makers are placed in a scene and the camera is used to photograph them from a range of locations [6]. The transformation the camera undergoes between successive posses is then calculated. This is combined with the known transformations the robot's hand undergoes to give the transformation between them. More recent approaches to hand-eye calibration have made use of structure from motion techniques to allow them to operate without requiring markers or chequerboards [7].

II. Method

A. Notation

The following notation will be used throughout this paper

T_{VC}	The transformation from the Velodyne lidar to the camera.
T_{Vi}	The transformation from the Velodyne's location at timestep i-1 to its location at timestep i
T_{Ci}	The transformation from the camera's location at timestep i-1 to its location at timestep i
S	Constant that accounts for the scale ambiguity of the camera transformation.
R	Rotation matrix
t	translation vector

B. Overview

Our approach can be roughly divided into four stages. The calculation of T_{Vi} and T_{Ci}, the estimation of R_{VC}, the estimation of t_{VC} and the refinement of T_{VC}. The method used in each instance is outlined below.

C. Calculation of T_{Vi} and T_{Ci}

The set of transforms T_{Ci} that describe the movement of the camera are calculated, up to scale ambiguity using a structure from motion approach. To calculate T_{Vi}, the set of transforms for the Velodyne lidar, the *iterative closest point* (ICP) algorithm is used.

D. Estimation of R_{VC}

Equation 1 gives the relationship between the rotation matrices.

$$R_{Ci} = R_{VC} R_{Vi} \tag{1}$$

In our implementation two methods for obtaining R_{VC} are used and combined to obtain an estimate of the rotations. For the first method the rotation component of each of the sensors transform is used. A solution to equation 1 is found by first converting the sensors rotations to angle-axis representation, where a_i is the angle and A_i the axes vector. These vectors are related by equation 2

$$A_{Ci} = R_{VC} A_{Vi} \tag{2}$$

The least squared solution to this equation can be found using the Kabsch algorithm. To give the rotation and an estimate of the error a sub-sample of 10 points is selected and the Kabsch algorithm is used to find the error. This is performed 500 times. The solution giving the smallest median error over the whole dataset is taken and the remaining solutions used to calculate the standard deviation for the roll, pitch and yaw.

When used on robotic arms the above method gives accurate results, however when used on an autonomous vehicle on a typical drive significant error is present after this step. This is due to the highly constrained movement these vehicles undergo. The same restrictions on the motion, however enable the use of a second method for calculating the rotation. The relationship between the transforms is given by equation 3

$$R_{Ci} t_{VC} + S t_{Ci} = R_{VC} t_{Vi} + t_{VC} \tag{3}$$

If we assume $R_{Ci} \approx I$ as is the case in most of the frames recorded from the car then the above equation simplifies to equation 4

$$S t_{Ci} \approx R_{VC} t_{Vi} \tag{4}$$

If both sides of the equation are normalized, the scale ambiguity is removed and this becomes equation 5

$$t_{Ci} \approx R_{VC} t_{Vi} \tag{5}$$

An equation with the same form as equation 2, that can be solved via the same method. The two approaches outlined both have significant rotational error about at least one axis. These large errors however generally lies on different axes, meaning if the two results are combined one accurate solution is obtained. The methods are combined by assuming that the error is a Gaussian distribution.

E. Estimation of t_{VC}

Once the rotation matrix is known the translation of the sensors is fairly straight forward to calculate. By starting with equation 6.

$$R_{Ci} t_{VC} + S t_{Ci} = R_{VC} t_{Vi} + t_{VC} \tag{6}$$

The terms can be rearranged and combined with the equations of other timesteps to form the matrix equation below

$$\begin{bmatrix} R_{Ci} - I & t_{Ci} & 0 & 0 & \ldots \\ R_{C(i+1)} - I & 0 & t_{C(i+1)} & 0 & \ldots \\ R_{C(i+2)} - I & 0 & 0 & t_{C(i+2)} & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{bmatrix} \begin{bmatrix} t_{VC} \\ S_i \\ S_(i+1) \\ S_(i+2) \\ \ldots \end{bmatrix} = \begin{bmatrix} R_{VC} t_{Vi} \\ R_{VC} t_{V(i+1)} \\ R_{VC} t_{V(i+2)} \\ \ldots \end{bmatrix} \tag{7}$$

The only unknowns here are t_{VC} and the scale. The estimate for these values are calculated by subsampling the data in the same manner as for the rotation calculations.

F. Refining the estimate of T_{VC}

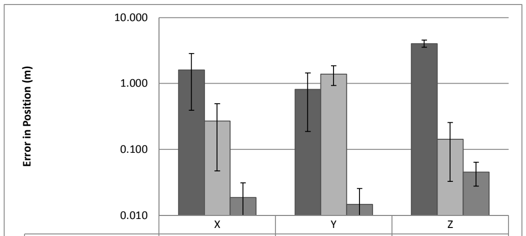
From the structure from motion methods an estimate of the rotation and translation between the sensors has been obtained as well as an estimate of the standard deviation of the error. From these starting conditions we make use of the metric proposed by Levinson et al [1]. The metric is optimized using a particle swarm optimization.

Particle swarm optimisation works by using a group of particles to explore the search space. We set the initial location of the particles by using the probability distributions obtained previously. The roll, pitch, yaw, x, y and z values are randomly drawn from a Gaussian distribution with the mean and standard deviation given by the hand-eye calibration step. The x,y and z position estimates are also constrained to be within 5m of the origin. A 5m limit on the offset between the sensors ensures that configurations that would be physically impossible for a typical mobile sensor platform are not evaluated.

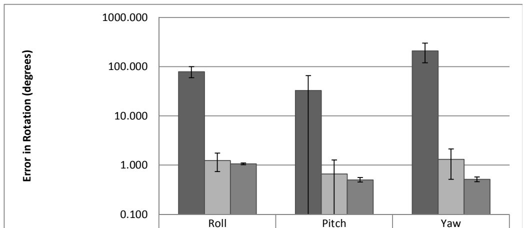
By shaping the initial positions of the particles to represent the calibration estimate and its confidence from the hand eye calibration we greatly reduce the search space that must be explored for the selected metric to find the correct calibration.

III. Results

All of the results presented here were tested using drive 28 of the KITTI dataset [8]. The vehicle used to collect this dataset has two sets of forward facing stereo cameras, a Velodyne lidar and a gps. In all our tests only the Velodyne and left most monochrome camera was used.



	X	Y	Z
Unconstrained Levinson Method	1.618	0.820	4.053
Hand-eye	0.270	1.397	0.144
Combined approach	0.019	0.015	0.046



	Roll	Pitch	Yaw
Unconstrained Levinson Method	79.501	32.881	209.554
Hand-eye	1.247	0.665	1.322
Combined approach	1.064	0.506	0.517

TABLE I

ACCURACY COMPARISON OF METHODS ON KITTI DATASET. NOTE TABLE USES LOG SCALE.

The overall calibration process was tested by randomly selecting a range of 500 frames representing 50 seconds of data from the dataset. The hand-eye calibration was first performed on all 500 frames, before the transformation was refined using Levinson's method on every tenth frame. The experiment was repeated 100 times with the mean and standard diviation of the results recorded. Levinson's method was chosen because it was developed for Velodyne data and has been shown to provide the best results for this type of sensors.

The experiment was also performed using the Levinson method without first constraining the initial conditions of the particle swarm optimizer using the hand-eye calibration method. Instead the initial population was distributed randomly so that the translation was within 3m and all angles were possible. This was done so that the impact our method has can be clearly seen. The results of this experiment are shown in Table I.

In all cases the unconstrained Levinson method performed poorly. This was to be expected as the optimization was performed over a large 6-dof search space with only 200 points evaluated on each iteration. The Hand-eye calibration on the other hand performed well on the rotations with an average error of around 1 degree and poorly on the translations with an error of around 1m. Again this was to be expected as the small rotations a typical vehicle undergoes while moving result in the translation estimation step of hand-eye calibration being extremely sensitive to noise. The combined approach dramatically improved upon the translation estimate of the other two solutions as well as slightly improving the rotation estimate provided by the Hand-eye calibration solution. In this method due to the accurate estimates of rotation provided by the hand-eye calibration the Levinson method was only required to search a very small sliver of the rotation search space resulting in the 200 points evaluated each iteration being sufficient to find the global minimum. An example of the output of our approach is shown in Figure 1.

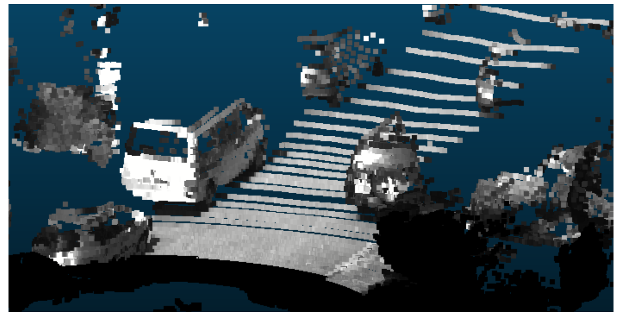


Fig. 1. Example of camera image projected onto lidar after alignment

IV. CONCLUSION

A new method for calibrating the extrinsic properties of a lidar-camera rig mounted to a ground vehicle has been developed. The method combines the strengths of standard automatic lidar-camera registration techniques with hand-eye calibration techniques developed for robotic arm calibration. The resulting method does not require the user to have any knowledge of the sensors layout or their internal co-ordinate frames while still achieving accurate calibrations.

ACKNOWLEDGEMENTS

This work has been supported by the Rio Tinto Centre for Mine Automation and the Australian Centre for Field Robotics, University of Sydney.

REFERENCES

[1] J. Levinson and S. Thrun, "Automatic Calibration of Cameras and Lasers in Arbitrary Scenes," in *International Symposium on Experimental Robotics*, 2012.

[2] G. Pandey, J. R. Mcbride, S. Savarese, and R. M. Eustice, "Automatic Targetless Extrinsic Calibration of a 3D Lidar and Camera by Maximizing Mutual Information," *Twenty-Sixth AAAI Conference on Artificial Intelligence*, pp. 2053–2059, 2012.

[3] Z. Taylor and J. Nieto, "A Mutual Information Approach to Automatic Calibration of Camera and Lidar in Natural Environments," in *the Australian Conference on Robotics and Automation (ACRA)*, 2012, pp. 3–5. [Online]. Available: http://www.araa.asn.au/acra/acra2012/papers/pap113.pdf

[4] A. Napier, P. Corke, and P. Newman, "Cross-Calibration of Push-Broom 2D LIDARs and Cameras In Natural Scenes," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2013.

[5] Z. Taylor, J. Nieto, and D. Johnson, "Automatic calibration of multi-modal sensor systems using a gradient orientation measure," *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1293–1300, 2013.

[6] R. Time, V. Robotics, and E. Calibration, "The New Approach Bmic Setup," pp. 554–561, 1988.

[7] N. Andreff, "Robot Hand-Eye Calibration Using Structure-from-Motion," *The International Journal of Robotics Research*, vol. 20, no. 3, pp. 228–248, Mar. 2001.

[8] A. Geiger and P. Lenz, "Are we ready for autonomous driving? the kitti vision benchmark suite," *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3354–3361, 2012.