Cross-Calibration of Push-Broom 2D LIDARs and Cameras In Natural Scenes

Ashley Napier* and Peter Corke** and Paul Newman*

***Abstract*— This paper addresses the problem of automatically estimating the relative pose between a push-broom LIDAR and a camera without the need for artificial calibration targets or other human intervention. Further we do not require the sensors to have an overlapping field of view, it is enough that they observe the same scene but at different times from a moving platform. Matching between sensor modalities is achieved without feature extraction. We present results from field trials which suggest that this new approach achieves an extrinsic calibration accuracy of millimeters in translation and deci-degrees in rotation.**

I. Introduction

Two of the most prevalent sensors in robotics, especially in the transport domain, are cameras and LIDAR (Light Detection And Ranging). Increasingly these sensors are being used to supplement each other, an examples of which is Google Street View, here LIDARs are used in conjunction with an omnidirectional camera to display planes in the captured images [1]. A Multi-modal approach to localisation was also presented in [11], here a monocular camera is localised using scene geometry provided by a 3D LIDAR sensor. In order to achieve this projection of LIDAR scans into the images taken by a camera, the extrinsic calibration between the various sensors must be known to a high degree of accuracy [5].

This extrinsic calibration of sensors can be performed in many different ways. The simplest approach is to physically measure the position of sensors relative to each other. This approach, however, proves to be more complicated than first thought as sensors are generally housed in casings which don't allow for accurate measurement of the sensing element itself.

Another approach is to place calibration targets into the workspace which are simultaneously in the field of view of both sensors. Calibration is performed by aligning features on the calibration targets observed by both sensors. A method of extrinsic calibration of a camera with a 2D range finder using a checkerboard pattern was presented in [12]. Checkerboards are again used in [6] and [2] to cross-calibrate 3D LIDAR with a camera requiring varying amounts of user guided preprocessing.

Robust long-term autonomy, however, requires a continuous assessment of calibration accuracy, which makes the use of calibration targets impractical. The sensors considered here are often included in safety critical systems so the ability to test the validity of calibrations or recalibrate after bumps, knocks and vibrations is critical to reliable operation. A method of target-less extrinsic calibration, which required the user to specify several point correspondences between 3D range data and a camera image was presented in [8]. Levinson [4] reduced the need for user intervention by examining edges in images, from an omnidirectional camera, which are assumed to correspond to range discontinuities in 3D LIDAR scans. Our approach is most closely related to a method described by [7], where reflectance values from a static 3D LIDAR are registered against captured images using an objective function based on Mutual Information.

Fig. 1. An example of a swathe which can be thought of as a 3D point cloud built up as a push-broom 2D LIDAR sweeps through an environment with vehicle motion. The red lines illustrate the latest scan from the 2D LIDAR, which is mounted on the front of our Bowler WildCat research platform. This swathe was constructed using calibration estimates obtained from the proposed method relative to a stereo camera providing a trajectory derived from visual odometry.

In contrast to prior art, our approach does not require both sensors to have overlapping fields of view. Instead, we exploit the motion of the vehicle to retrospectively compare LIDAR data with camera data. This is achieved through the generation of a swathe of LIDAR data built up as the LIDAR mounted in a push-broom configuration traverses through the workspace, figure 1. Therefore, our method can be used as long as the motion of the vehicle causes eventual overlap of the observed workspaces. This feature of the calibration is particularly useful in transport as often sensors can be mounted all over the vehicle with non-overlapping fields of view. It should also be noted that 2D LIDAR are currently cheaper than 3D LIDAR by a couple orders of magnitude

*Mobile Robotics Group, University of Oxford.
{ashley,pnewman}@robots.ox.ac.uk
**CyPhy Lab, School of Electrical Engineering & Computer Science, Queensland University of Technology, Brisbane, Australia.
peter.corke@qut.edu.au

and much easier to mount discretely making them a much more attractive prospect for use in commercial autonomous vehicles.

We pose the calibration as a view matching problem and require no explicit calibration targets or human intervention, as has been required in most prior art [12], [7], [2]. Instead we explicitly exploit the fact that scenes in laser light look similar to scenes as recorded by an off-the-shelf camera. We synthesise images from LIDAR reflectance values based on the putative calibration between sensors and measure how well they align, figure 3. Rather than use a feature based approach to measure alignment we exploit the gross appearance of the scene using a robust metric in the form of a gradient based Sum of Squares objective function. The calibration giving maximal alignment is then accepted to be the best estimate of the camera LIDAR calibration. As far as we are aware this is the first piece of work to present automatic calibration of a camera and 2D LIDAR in natural scenes without explicit targets placed into the workspace or other user intervention. This is also the only piece of work not requiring sensors to be mounted such that they have overlapping fields of view.

II. Problem Formulation

In order to create a swathe with a 2D push-broom LIDAR it must undergo motion through its environment. Specifically, we construct a swathe using a base trajectory estimate, $X^b(t)$, obtained using an INS or, in our case, visual odometry and the putative calibration ${}^bT_l$ between the base trajectory and the LIDAR l. The swathe is then projected into the camera using the current calibration between the camera c and base trajectory ${}^bT_c$, figure 2. An interpolated LIDAR reflectance image is then generated, figure 3. We use an edge-based, weighted SSD (Sum of Squares Distance) objective function to measure the alignment of an image captured by the camera and the LIDAR reflectance image. A simple iterative optimisation is used to search over the $\mathbb{SE}(3)$ pose which defines the extrinsic calibration and maximises the alignment of the camera image and the generated LIDAR reflectance image. The best estimate of the extrinsic calibration achieves the best alignment of the two images.

A. Generating A Swathe

To generate a metrically correct swathe from the push-broom LIDAR requires accurate knowledge of the sensor's motion. In the general case shown in Figure 2, a base trajectory $X^b(t)$ is a full $\mathbb{SE}3$ pose, $(x, y, z, roll, pitch, yaw)$, as a function of time. $X^b(t)$ can be derived from a multitude of sensors including inertial navigation systems (INS) and visual odometry (VO), as long as the trajectory is metrically accurate over the scale of the swathe. The poses of the LIDAR and camera are given relative to this base trajectory. In order to ensure millisecond timing accuracy we use the TicSync library [3] to synchronise the clocks of the sensors to the main computers.

Let the i^{th} LIDAR scan be recorded at time t_i and consist of a set of points, $\mathbf{x_i}$, and a set of corresponding

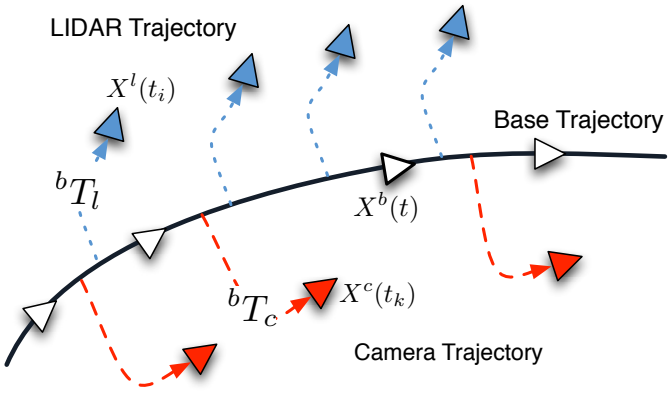

Fig. 2. Shows the base trajectory $X^b(t)$, the camera $X^c(t_k)$ and LIDAR $X^l(t_i)$ trajectories relative to it. It should be noted that this is the general case, in the results presented here $X^b(t)$ is generated using visual odometry so the camera trajectory and base trajectory are one and the same reducing the dimentionality of the search space from twelve degrees of freedom to six.

reflectance values, $\mathbf{R_i}$, such that laser point j in this scan, $x_{ij} = [x_j, y_j]^T$, is associated with reflectance value $R_{i,j}$. We currently make the approximation that all points j within the ith scan are captured at the same time, in reality each scan takes $20ms$. As the data used for calibration was collected at purposefully slow speeds this approximation has a negligible effect. We first compute the pose of the LIDAR $X^l(t_i)$ based on the current putative extrinsic calibration ${}^bT_l$ and the base trajectory.

$$X^l(t_i) = X^b(t_i) \oplus {}^bT_l \tag{1}$$

Where $\oplus$ denotes a composition operator. Each scan can then be projected into a local 3D scene $\mathbf{P_i}$ creating a swathe of laser data.

$$\mathbf{P_i} = X^l(t_i) \oplus \mathbf{x_i} \tag{2}$$

We can now generate a swathe as a function of the extrinsic calibration between the sensor base trajectory and the LIDAR. An example is shown in figure 1.

B. Generating LIDAR Reflectance Imagery

The next stage is generating LIDAR reflectance images as viewed from the pose of the camera c capturing an image $\mathcal{I}_k^c$ at time t_k. First, the swathe $\mathbf{P_i}$ is transformed into the camera's frame of reference using the current estimate of the extrinsic calibration between the base trajectory and the camera, ${}^bT_c$. The pose of the camera $X^c(t_k)$ at time t_k is then written as

$$X^c(t_k) = X^b(t_k) \oplus {}^bT_c \tag{3}$$

The swathe is then transformed into the camera's frame and projected into the camera using the camera's intrinsics, K, which are assumed to be known. Thus,

$$\mathbf{p_{i,k}} = proj(\ominus X(t_k) \oplus \mathbf{P_i}, K) \tag{4}$$

gives us the pixel locations of the swathe points $\mathbf{p_{i,k}}$ in the camera image $\mathcal{I}_k^c$. At this point we could use the individual LIDAR reflectance values $\mathbf{R_{i,k}}$ and compare their reflectivity

a) b) c)

Fig. 3. a) An interpolated laser reflectance image at the estimated extrinsic calibration $\mathcal{I}_k^l({}^cT_l)$ created from LIDAR reflectance values projected into the camera. b) The actual camera image $\mathcal{I}_k^c$. c) An interpolated laser image with incorrect calibration $\mathcal{I}_k^l({}^cT_{0l})$. This scene was observed during a parking maneuver.

to the pixel intensities $\mathcal{I}_k^c(\mathbf{p_{i,k}})$. However, the density of the points is extremely variable due to foreshortening effects as points at larger ranges from the camera map to a smaller areas within the image. We therefore use cubic interpolation to sample the intensities $\mathbf{R_{i,k}}$ at pixel locations $\mathbf{p_{i,k}}$ over the same grid as the pixels in $\mathcal{I}_k^c$. This generates a laser reflectance image $\mathcal{I}_k^l({}^bT_c, {}^bT_l)$ as a function of the extrinsic calibration, an example of which can be seen in Figure 3. In the results presented in this papers the base trajectory $X^b(t)$ is derived from stereo visual odometry [10]. This simplifies the extrinsic calibration as the base frame is equivalent to the camera frame reducing ${}^bT_c$ to the identity and, in turn, the search space from twelve degrees of freedom to six. The laser reflectance image then becomes a function only of ${}^bT_l$, which is equivalent to ${}^cT_l$, the extrinsic calibration between the LIDAR and camera.

C. The Objective Function

At this point we have the ability to take a single camera image $\mathcal{I}_k^c$ and generate, given data from a 2D LIDAR and knowledge of the platform trajectory, a corresponding laser reflectance image $\mathcal{I}_k^l({}^cT_l)$ based on a putative extrinsic calibration between the two sensors. We now seek a metric which accurately reflects the quality of the alignment between the two images. This task is made difficult by non-linearities in the reflectance data [9] rendering basic correlation measures such as mutual information and standard SSD ineffective. It was found empirically that taking a smoothed gradient image was far more stable. Further, patch-based normalisation is applied whereby local variations in gradient are normalised to be consistent across the whole image or at least between corresponding patches in $\mathcal{I}_k^c$ and $\mathcal{I}_k^l({}^cT_l)$. Applying patch based normalisation enhances local image gradients and avoids very strong edges completely dominating the objective function, see Figure 4. The pixel values from both images are then weighted by $w_{\mathcal{I}_k^c}$ the inverse of the distance transform of the reflectance measurement $\mathbf{p_{i,k}}$ over the image grid, giving extra weight to areas with a higher sampling density. The objective function can thus be expressed as

$$O({}^cT_l) = \sum_{\mathcal{I}_k^c} w_{\mathcal{I}_k^c} \left\| Q(\mathcal{I}_k^c) - Q(\mathcal{I}_k^l({}^cT_l)) \right\|_2 \qquad (5)$$

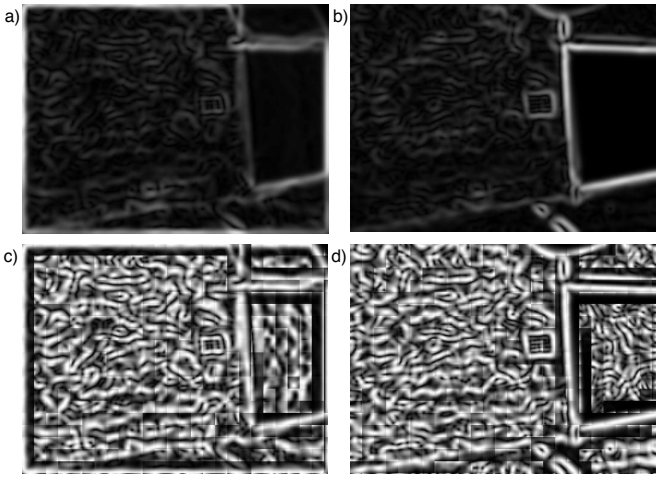

Fig. 4. The smoothed edge image from the laser a) and the corresponding camera image b) from Figure 3. c) and d) are the same gradient images after the patch based normalisation procedure respectively. Note how the details are emphasised by the patch-based normalisation leveraging details that would have been drowned out by the few dominant edges if the sum of squares distance of a) and b) were computed directly.

where $\sum_{\mathcal{I}_k^c}$ denotes the sum over all pixels in the image pair $\mathcal{I}_k^c$ and $\mathcal{I}_k^l({}^cT_l)$ and $Q(\bullet)$ denotes a function which performs Gaussian smoothing before taking the magnitude gradient image and performing patch based normalisation. In the results presented here we use a Gaussian kernel of 25x25 pixels with a variance of 6.5 and a patch size of 20x20 pixels for the patch based normalisation procedure.

D. Optimisation

The objective function provides a pronounced narrow convergence basin around the correct solution, see figure 9. We therefore use a simple iterative optimisation to find a solution. As this calibration is not required to be a realtime application (it can be run as a background process on a vehicle), high speed is not a priority. Therefore, starting from an initial estimate ${}^cT_{lo}$, the search for a minimum is conducted along each degree of freedom individually, updating the estimate of ${}^cT_l$ as it proceeds. For the results presented in this paper we used a range of 30cm and 10

degrees with a resolution of ~1mm and 0.02 degrees. While this brute force optimisation method was found to work well in our experiments our approach is agnostic to the exact optimisation method used. The estimate for a calibration for a particular image at time k is explicitly written as

$$ {}^{c}\bar{T}_l = \underset{{}^{c}T_l}{argmin} \sum_{\mathcal{I}_k^c} w_{\mathcal{I}_k^c} \left\| Q(\mathcal{I}_k^c) - Q(\mathcal{I}_k^l({}^{c}T_l)) \right\|_2 \quad (6) $$

E. Improving Accuracy With Estimate Fusion

We can now obtain, at any time, an estimate for the calibration between the LIDAR and camera. What remains to be asked is

a) are all scenes equally useful in supporting the cross model calibration?
b) how might we fuse multiple calibration estimates?

The answer to these questions are one in the same. Looking at figure 9 we gain the intuition that minima at the bottom of a sharp trench are more informative than those at the bottom of a rough, broad bowl. By flipping the cost function and fitting a Gaussian $G({}^{c}\bar{T}_l, \sigma^2)$ to the the peaks (which were minima) we can obtain a likelihood function $\mathcal{L}({}^{c}T_l)$. This is the Laplace approximation and results in a "measurement" model parameterised by $\sigma^{2\cdot}$.

We are now in a position to fuse a sequence of noisy measurements of the latent state ${}^{c}T_l$. Such a sequence ${}^{c}\mathcal{T}_l = ({}^{c}\bar{T}_{l1}, {}^{c}\bar{T}_{l2}, {}^{c}\bar{T}_{l3}..., {}^{c}\bar{T}_{lN},)$ with associated variances $\sum = (\sigma_1^2, \sigma_2^2, \sigma_3^2, ..., \sigma_N^2,)$ is fused via a recursive Bayes filter, allowing us to sequentially update our calibration as new parts of the workspace are entered, without the need for an expensive joint optimisation over N frames. This process of treating each optimisation as a noisy measurement significantly reduces the standard deviation of the calibration result when compared to the deviation of the raw estimates from individual frames, see figure 6 and table I.

III. Experimental Results

We used our Wildcat platform with a Point Grey bumblebee2 stereo camera and SICK LMS-100 2D LIDAR, see figure 5, to evaluate the performance of the proposed algorithm. The results shown here were obtained using several scenes from around our field center at Begbroke. Swathes of approximately 300 LIDAR scans, equaling over 140,000 measurements, were used for each individual calibration estimate. 73 different images from the left camera of the stereo pair were used to validate the calibration estimate.

Trajectory estimates were provided by our in house stereo visual odometery (VO) system [10]. Obtaining ground truth for trajectory estimation is alway problematic, however the VO system has demonstrated high local metric accuracy, drifting as little as 3 meters over a 700 meter closed loop trajectory.

A. Evaluating The Performance Of Estimate Fusion

In order to estimate the performance and repeatability of the proposed algorithm we first evaluated the calibration for

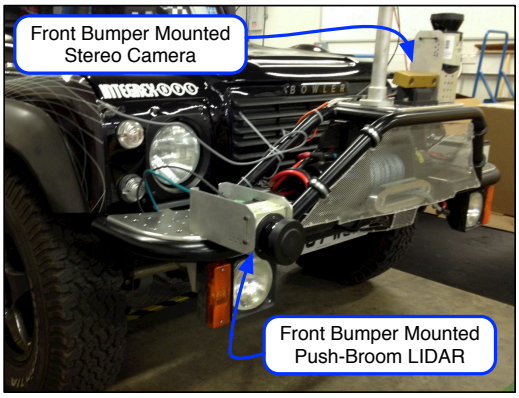

Fig. 5. The WildCat platform used in our experiments is a fully equipped autonomous vehicle, it has onboard computing and multiple sensors including a SICK 2D LIDAR and a PointGrey BumbleBee stereo camera. The platform enables fast and reliable collection of survey quality data used in this work. It should also be noted that the 2D LIDAR and stereo camera do not have overlapping fields of view when stationary, meaning a retrospective calibration technique as presented here is required for data-based extrinsic calibration.

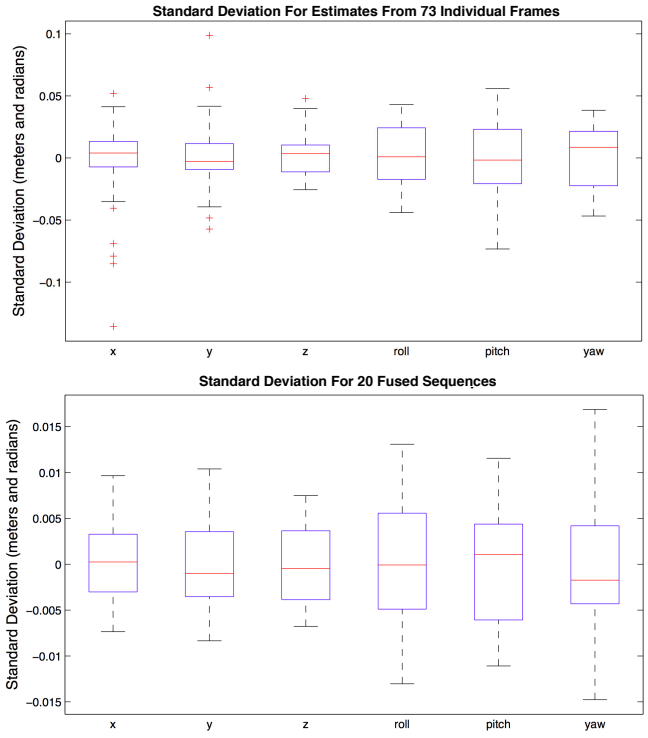

Fig. 6. Box plots of the standard deviation of the extrinsic calibration estimate in each of the six degrees of freedom. The top plot shows the standard deviation of optimisation results from single frames while the bottom shows the deviation after fusing multiple frames using Laplace's approximation and a recursive Bayes filter. Twenty fused results were generated by selecting sequences of ten random calibration estimates from the possible 73 single frames in a M pick N trial. It can be clearly seen that after fusion the standard deviation is significantly reduced, see table I for details, and outliers are effectively ignored. Boxes extend from the 25th to 75th percentiles, whiskers extent to the extrema of the data not considered to be outliers, which are denoted by red crosses. Note the different vertical scales.

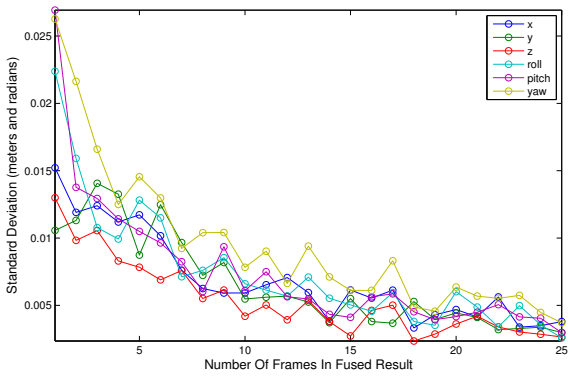

Fig. 7. A plot showing how the standard deviation of the calibration estimate reduces as more frames are encountered and the calibration estimate is updated. This data represents twenty sequences of 25 frames randomly selected from a possible 73 frames.

Standard Deviation	***Translation (mm)***			***Rotation (degrees)***		
	x	y	z	roll	pitch	yaw
Individual Results	28	21	15	1.4	1.4	1.5
Fused Results	4.5	5.2	4.6	0.38	0.39	0.44

TABLE I

TABLE SHOWS THE STANDARD DEVIATION OF THE CROSS-CALIBRATION ESTIMATE OVER 73 FRAMES FROM SEVERAL SCENES, TEN OF WHICH ARE SHOWN IN FIGURE 8. THE EFFECT OF FUSING ESTIMATES USING THE BAYES FILTER CAN BE CLEARLY SEEN WITH ALMOST AN ORDER OF MAGNITUDE REDUCTION IN THE STANDARD DEVIATION OF THE CALIBRATION ESTIMATES.

73 individual frames. The results for this experiment can be seen in figure 6 and table I. The standard deviation of the estimates — which is important as we are unable to accurately measure ground truth — for individual frames is of the order of a couple of centimeters and degrees, which is akin to the expected level of accuracy a hand measured calibration would yield. In order to test any improvement achieved by fusion of individual frames we performed twenty N choose M trials with N being the 73 frames and $M = 10$. For each fused estimate ten random frames were chosen and then fused in sequence using the Bayes filter. An example set of ten frames can be seen in figure 8. The effect of the fusion stage can be seen in figure 6 and table I with the standard deviations decreasing by almost an order of magnitude.

Figure 7 shows how the standard deviation of calibration estimates decrease as more individual estimates are fused together, the effect appears to saturate after approximately ten frames.

Figure 9 shows how frames with ill conditioned and noisy calibration estimates, see figure $9(b)$, are automatically assigned a higher σ^2, as per the process in Section II-E. Estimates with more distinct troughs around the minima, figure $9(a)$, conversely are assigned a lower σ^2 and as expected contribute more information to the final calibration estimate after fusion. This is illustrated by the bar charts in figure 9, which plots the inverse variances, $1/\sigma^2$, which effectively weight the individual estimates in the Bayes filter. Here the frame in figure $9(a)$ is given over 50% more weight that the frame in figure $9(b)$.

Given that the stereo camera and 2D LIDAR do not have instantaneously overlapping fields of view this calibration would not be possible with any of the other techniques reported in the literature.

B. Objective Function Around Solution

Figure 9 shows the objective function plotted about the estimated extrinsic calibration for two frames. It can be seen that there is a distinct convex peak in a) but less so in b). However, away from the true solution the objective function is non convex, which justifies the choice of the simple search based optimisation over a traditional gradient-based method. It is possible that the final calibration could be improved by a final gradient decent type optimisation once the search has found a coarse solution, this has not been investigated in this work. It can also be seen that some degrees of freedom are more sensitive than others, this is thought to be due to the geometry of the typical scene and sensor placement. Note this is handled naturally by our filtering.

IV. Conclusions and Future Work

We have presented an automatic calibration procedure for a camera and a 2D LIDAR under general motion. A method which can be used in natural scenes without the need for targets, enabling on the fly calibration. The method also does not require sensors to be mounted such that they have overlapping views. Unlike other approaches we exploit the motion of the vehicle using a trajectory estimate to build a swathe of LIDAR data to generate laser reflectance imagery for comparison to images captured by the camera. We have adopted a robust correlation measure that is invariant to non-linearities in the reflectance returns and camera images. Furthermore we have demonstrated the calibration leveraging exposure to multiple scenes and fusing multiple one-shot calibrations yielding accuracies of millimeters in translation and deci-degrees in rotation. This is done in a way which is sympathetic to the utility of the scene structure and appearance in constraining the extrinsic calibration parameters.

While these results are compelling there remains much to be done in our future work. We intend to investigate the effect of lighting conditions on the procedure, extend the optimisation to cover multiple lasers and perhaps most interestingly, use it to register multiple 2D lasers to each other without recourse to a camera.

V. Acknowledgments

The work in this paper was funded by an EPSRC DTA. Paul Newman is supported by the EPSRC Leadership Fellowship EP/J012017/1. Peter Corke was supported by Australian Research Council project DP110103006 Lifelong Robotic Navigation using Visual Perception. The authors gratefully acknowledge the support of BAE Systems who provided the WildCat platform.

Fig. 8. A set of 10 randomly selected frames used in a fused calibration estimate. It can be seen that the frames are at acquired from various positions within different scenes

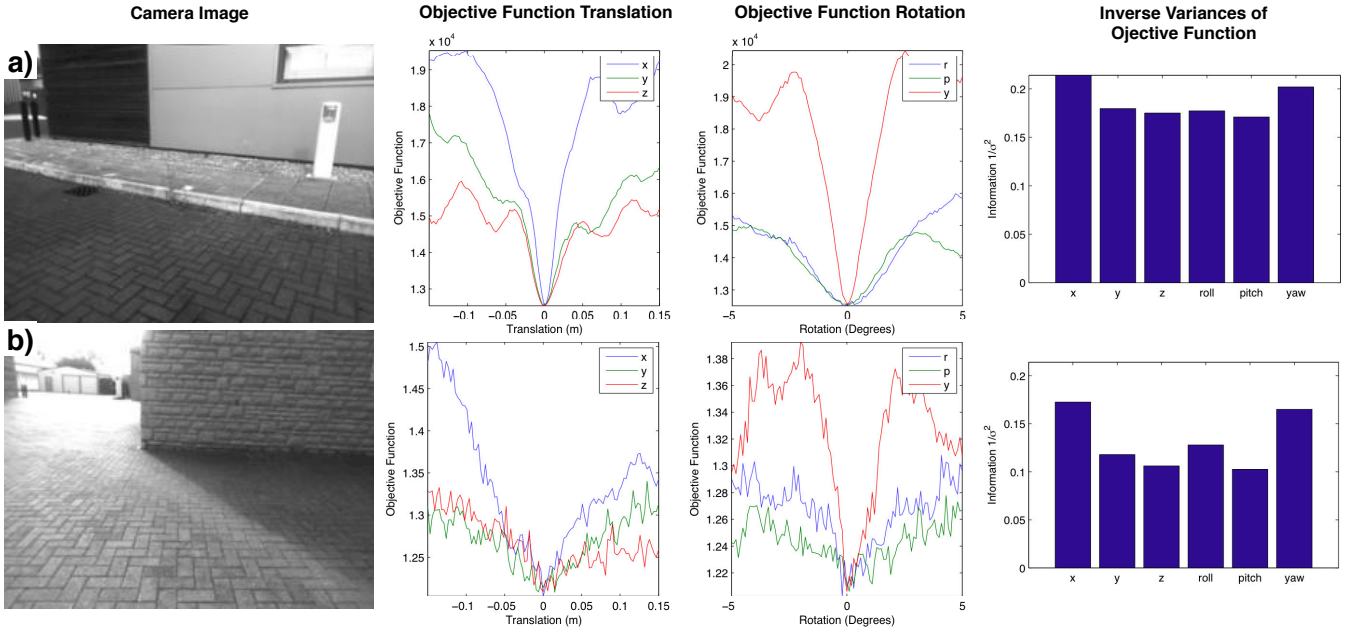

Fig. 9. The objective function plotted about the estimated calibration for two example scenes. a) shows an example of how the objective function is convex around the solution and has a clear global minima. Away from the minima the objective function can be non convex, justifying the use of the grid search approach improving the basin of convergence and reducing the required accuracy of the initialization. However, b) shows how certain scenes can produce noisy and ill conditioned objective functions around the solution. The bar charts of inverse variance of the estimate derived from the shape of the objective function at the minima demonstrate how less pronounced minima yield less information and are given less weight in the Bayes filter.

REFERENCES

[1] Dragomir Anguelov et al. Google street view: Capturing the world at street level. *Computer*, 43:32–38, 2010.

[2] Andreas Geiger, Frank Moosmann, Omer Car, and Bernhard Schuster. Automatic camera and range sensor calibration using a single shot. *IEEE Int. Conf. on Robotics and Automation*, 2012.

[3] Alastair Harrison and Paul Newman. Ticsync: Knowing when things happened. *Proc. IEEE International Conference on Robotics and Automation*, 2011.

[4] Levinson Jesse and Sebastian Thrun. Automatic calibration of cameras and lasers in arbitrary scenes. *International Symposium on Experimental Robotics*, 2012.

[5] Quoc V. Le and Andrew Y. Ng. Joint calibration of multiple sensors. *Intelligent Robots and Systems*, 2009.

[6] Gaurav Pandey, James McBride, Silvio Savarese, and Ryan Eustice. Extrinsic calibration of a 3d laser scanner and an omnidirectional camera. *Intelligent Autonomous Vehicles*, 2010.

[7] Gaurav Pandey, James R. McBride, Silvio Savarese, and Ryan M. Eustice. Automatic targetless extrinsic calibration of a 3d lidar and camera by maximizing mutual information. *In Proceedings of the AAAI National Conference on Artificial Intelligence*, 2012.

[8] D. Scaramuzza, A. Harati, and R. Siegwart. Extrinsic self calibration of a camera and a 3d laser range finder from natural scenes. *Intelligent Robots and Systems, IROS. IEEE/RSJ International Conference on*, 2007.

[9] Ahmed Shaker, Wai Yeung Yan, and Nagwa El-Ashmawy. The effects of laser reflection angle on radiometric correction of the airborne lidar inensity data. *International Society for Photogrammetry and Remote Sensing*, 2011.

[10] G Sibley, C Mei, I Reid, and P Newman. Vast-scale Outdoor Navigation Using Adaptive Relative Bundle Adjustment. *The International Journal of Robotics Research*, 2010.

[11] Alex Stewart and Paul Newman. Laps - localisation using appearance of prior structure: 6-dof monocular camera localisation using prior pointclouds. *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, May 2012.

[12] Qilong Zhang and R. Pless. Extrinsic calibration of a camera and laser range finder. *Intelligent Robots and Systems (IROS). Proceedings. IEEE/RSJ International Conference on*, 2004.