Automatic Calibration of Lidar and Camera Images using Normalized Mutual Information

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

***Abstract*—This paper is about automatic calibration of a camera-lidar system. The method presented is designed to be as general as possible allowing it to be used in a large range of systems and applications. The approach uses normalized mutual information to compare camera images with lidar scans of the same area. A camera model that takes into account orientation, location and focal length is used to create a 2D lidar image, with the intensity of the pixels representing a feature of the lidar scan that is chosen depending on the application. Particle swarm optimization is used to find the optimal model parameters. The method presented is successfully validated on a variety of cameras, lidars and locations, including scans of both urban and natural environments.**

I. Introduction

This paper looks at a method for automatically aligning a camera and a lidar scanner using scans of an arbitrary environment. The method estimates the extrinsic and some intrinsic parameters of the camera in an automated process with minimal human interaction. This process is designed to work for a large variety of laser scanners and cameras used in a range of environments.

Accurate calibration between lidar scanners and cameras is important as it allows each point in the cloud produced by the scanner to have a colour associated with it. These coloured points can then be used to build up richer models of the area. This calibration is quite challenging due to the very different modalities of the sensors and the nature of the output information. Due to the difficulty of aligning the sensors the majority of these systems are calibrated by hand. This is currently done using reflective markers, chequerboards or by painstakingly hand labelling large numbers of points. These methods are slow, labour intensive and often produce results with significant errors.

Automatic calibration is important for mobile robots relying in multi sensor modalities. The most current common approach is to perform calibration once using manual methods and assume this calibration remains unchanged for a period of time. In practice however, the calibration is rapidly degraded due to the robot motion, particularly for mobile robots working in rough environments such as mining trucks. Robust automatic calibration methods would allow robots to work in rough environments for longer periods of time.

Several automated methods for aligning the cameras do exist, however, all of these methods are only designed to work for a fairly small range of situations and lidar types. The lidar sensors these methods work with can be roughly divided into three main categories: i) Terrestrial systems that make a single high resolution scan of an area from a fixed location, ii) airborne systems that perform a similar task to terrestrial systems from the air and iii) mobile systems, such as the velodyne that are used to make a large number of very coarse scans of an environment and are typically used for navigation. There are also two different types of cameras that are often used in conjunction with these scanners, regular colour and multi- and hyperspectral cameras. Hyperspectral cameras have hundreds of colour bands and operate by scanning a single line of pixels over an area.

The attempt in this paper is to create a method that can correctly find the calibration for a lidar-camera system that can be used regardless of the exact type of sensors involved. The method also attempts to successfully operate in both urban and natural environments.

II. Related work

The work done in this area can be roughly divided into three groups, calibration of aerial scans, fixed ground based scans and mobile ground scans.

In the aerial scans application a recently proposed method by H. Li *et al* makes use of edges and corners [1]. Their method works by constructing closed polygons from edges detected in both the lidar scan and images. Once the polygons have been extracted they are used as features and matched to align the sensors. The method was only intended for and thus tested using aerial photos of urban environments.

A Mastin *et al* achieved registration of an aerial lidar scan by creating an image from it using a camera model [2]. The intensity of the pixels in the image generated from the lidar scan were either the intensity of the return the laser had or the height from the ground. The images were compared using mutual information and optimization was done via downhill simplex. This method operated quickly and produced accurate results although its search space was rather limited requiring an initial guess of the orientation of the camera that was correct to within 0.5 degrees for roll and pitch. The method was only tested in an urban environment where buildings provided a strong relationship between height and image colour.

For the alignment of fixed ground based scans in urban environments a large number of methods exist that exploit the detection of straight edges in a scene [3], [4]. These straight lines are used to calculate the location of vanishing points in the image. While these methods work well in cities and

with images of buildings they are unable to correctly register natural environments due to the lack of strong straight edges.

The registration of mobile ground scans is particularly challenging due to the low resolution of the lidar used. A recent approach that overcomes this issue has been presented by Levinson [5]. The method calibrated the extrinsic parameters of a camera-velodyne system using a set of 100 image scan pairs. Their method involves finding edge images for both the laser and camera images and using an element wise multiplication of these images, assuming that when the sum of this is maximized the two sensors are correctly aligned. There is also some extra processing done to improve the robustness and convergence of the method. A closely related work was recently presented [6]. The authors developed a method to calibrate a velodyne-camera system by maximizing the mutual information between laser reflectivity and optical images. There are two main differences with our system. First, our approach can be applied to a larger variety of environments since is not based on the solely use of reflectivity for the laser image. We found reflectivity to be uninformative in natural environments such as the mining data evaluated. The second difference regards the optimization method. In [6] the optimization is done by the Barzilai-Borwein (BB) steepest gradient ascent algorithm. We use particle swarm optimisation which is not restricted to convex problems and so allows us to perform single-scan calibration.

From a more theoretical view on the calibration [7] looked into different techniques for generating an image from a 3d model so that mutual information would successfully register the image with a real photo of the object. They used NEWUOA optimization in their registration and looked at using the silhouette, normals, specular map, ambient occlusion and combinations of these to create an image that would robustly be registered with the real image. They found surface normals and a combination of normal and ambient occlusion to be the most effective.

A fairly in depth look at many of the different methods for aligning images with lidar scans can be found in [8].

III. Method

Figure 1 shows a block diagram of our approach. The image is first converted to grey scale and histogram equalization is performed on the image. For the data given by the lidar, first the feature to use in colouring the image is chosen and the corresponding value for each point is generated. Once this had been done a camera model is used to create the 2D image from the point cloud. Normalized mutual information is used as a measure to compare the cameras image with the generated image. This process was repeated for changing camera model parameters using particle swarm optimization. The optimization continues until all the particles converge and a global maximum for the normalized mutual information between the images is found.

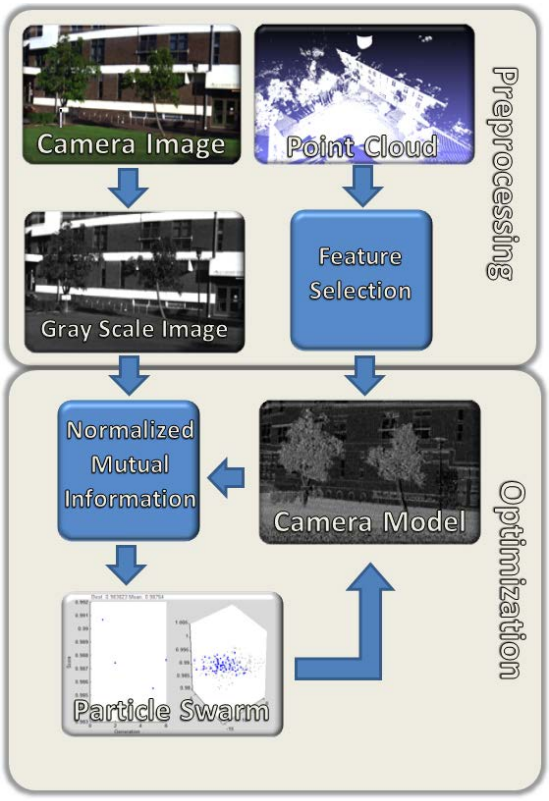

Fig. 1. Overview of alignment method

A. Lidar features

For the normalized mutual information (NMI) to correctly calibrate the system there has to be a strong statistical dependence between the colour of pixels in the camera image and the one generated for the lidar. The normals of the points and the return intensity of the lidar are used to achieve this. Normals are used as there is a fairly strong relationship between the angle of a surface and its perceived colour as was shown in [7].

Two different methods are used for estimating the normals. For the dense scans produced by the Riegl lidar, the normal vectors of the points are estimated by taking the difference between consecutive points in the scan. For sparse datasets, such as the velodyne this was found to give poor results and so a far more accurate, though slower method was implemented. A plane is approximated at the location of each point. This is done by first placing the points into a k-d tree. From this, the eight nearest neighbours to each point are found. The normal vector is calculated from the eigenvectors and eigenvalues of the covariance matrix C, given by equation 1 [9].

$$C = \frac{1}{8}\sum_{i=1}^{8}(p_i - c)(p_i - c)^T \tag{1}$$

Where p_i is the i-th nearest neighbour location and c is the location of the point. The smallest eigenvalue of C's corresponding eigenvector is the best estimate of the normal

vector of the plane. Once this vector has been obtained the angle between it and the horizontal $x-z$ plane is calculated and stored for use in the image generation process. The angle between the normals and a horizontal plane was used as it was assumed that most of the light was coming from above and thus this angle had the largest influence on intensity. Normals also have the advantage over many other methods of colouring the pixels in that they are independent of the location of the camera and so can be precalculated. This is important as the calculation of the image from the lidar scan is the most computationally expensive step in the optimization.

The intensity of the return of the rays in the lidar scan was also used as a feature. This gives a strong relationship as both laser reflectance and the camera pixel intensity primarily rely on the reflectance of the target material. It has the advantage, over any technique that makes use of the distance, in that it can detect a difference in the colour of an object. This allowed it to achieve the best registration on our second dataset (ACFR) due to aligning several lines that were painted on the side of a building. For the velodyne dataset the intensity of returns was of limited use as each laser appeared to give different readings for the same object.

A parameter that was also looked into was how areas where no lidar readings are obtained are treated. These areas, if included as features, could often enhance the registration accuracy. A strong example of this is the shape of the skyline in the lidar images. However, it could also lead to situations where it caused misregistration, for example by aligning the shadow cast by objects in the lidar with objects in the scene.

Histogram equalization is performed on all features to improve how they would be distributed into the bins during the mutual information calculation.

B. Mutual Information

Mutual information is a measure of how similar one signal is to another. It was first developed in information theory using the idea of Shannon entropy [10]. Shannon entropy is a measure of how much information is contained in a signal and its discrete version is defined as [11]:

$$H(X) = H(p_X) = \sum_{i=1}^{n} p_i log(\frac{1}{p_i}) \tag{2}$$

where X is a discrete random variable with n elements and the probability distribution $p_X = (p_1, ..., p_n)$. For this purpose $0log\infty = 0$.

Using this idea of Shannon entropy, mutual information is defined as

$$MI(M,N) = H(M) + H(N) - H(M,N) \tag{3}$$

where $H(M,N)$ is the joint entropy which is defined as

$$H(M,N) = H(p(m,n)) = \sum_{m}\sum_{n} p(m,n) log(\frac{1}{p(m,n)}) \tag{4}$$

Mutual information when used for registration purposes suffers from an issue in that it can be influenced by the amount of total information contained in images causing it to favour images with less overlap [12]. This is solved by using a normalized mutual information metric defined as

$$NMI(M,N) = \frac{H(M) + H(N)}{H(M,N)} \tag{5}$$

In practice, for images, the required probabilities $p(M)$ and $p(N)$ can be estimated using a histogram of the distribution of intensity values.

Normalized mutual information is used as the metric for evaluating the strength of the alignment between the two images as it automatically takes into account the non-linear relationship between angle and intensity. It also accounts for issues such as how different materials could appear dissimilar in different sensor modalities. This strength means that it can be assumed that the global maximum of normalized mutual information (NMI) is when the images are best aligned.

C. Camera model

To convert the lidar data from a list of 3D points to a 2D image that could be compared to the camera's images, the points are first pass into a transformation matrix that aligns the camera's and the world axis. After this has been performed, one of two basic camera models is used. For standard cameras a pin-hole camera model is used as defined in equations 6 and 7. For hyperspectral cameras, a panoramic camera model that projects the points onto a cylinder is used, a rough depiction of this is shown in figure 2. This model projects the points using equations 8 and 9 [13].

$$x_{cam} = x_0 - \frac{cx}{z} + \Delta x_{cam} \tag{6}$$

$$y_{cam} = y_0 - \frac{cy}{z} + \Delta y_{cam} \tag{7}$$

$$x_{cam} = x_0 - c\arctan(\frac{-y}{x}) + \Delta x_{cam} \tag{8}$$

$$y_{cam} = y_0 - \frac{cz}{\sqrt{x^2+y^2}} + \Delta y_{cam} \tag{9}$$

where

$\mathbf{x}_{cam}$, $\mathbf{y}_{cam}$ are the x and y position of the point in the image.

x, y, z are the coordinates of points in the environment.

c is the principle distance of the model.

$\mathbf{x}_0$, $\mathbf{y}_0$ are the location of the principle point in the image.

$\Delta\mathbf{x}$, $\Delta\mathbf{y}$ are the correction terms used to account for several imperfections in the camera.

These models ignore the effects of several other parameters such as the x and y axis of the camera not being perfectly parallel and the radial distortion of the lens. The ignoring of these effects can be justified as for a hyperspectral camera with a resolution of 10000 by 60000 it was shown in [13] that

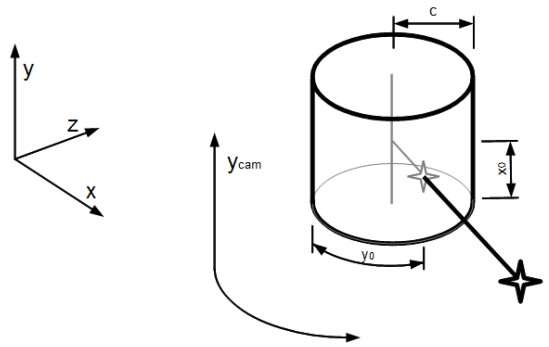

Fig. 2. Cylinder model used to represent hyperspectral camera

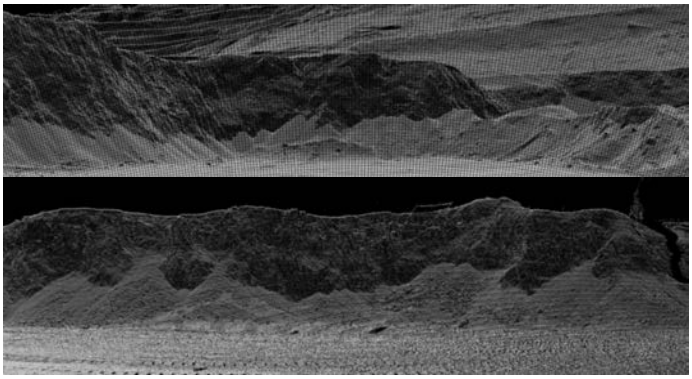

Fig. 3. Images of mining area output by the panoramic camera model

the error caused by these parameters was less than 10 pixels. This level of error in the images was taken to be acceptable. An example of the output of the camera model can be seen in figure 3.

D. Optimization

Depending on the assumptions made by the camera model and the accuracy of the initial scans position the problem has four to nine variables to solve. This search space is also highly non-convex with a large amount of local maximums. An example of the typical shape of NMI for a single scan alignment is plotted in two dimensions in figure 4. With the simple histogram method of calculating the mutual information used in this paper there is no information on the derivatives available. These difficulties were further compound by the relatively expensive process of generating an image from a pointcloud that is required for every function evaluation.

The fairly large range that the correct values could lie in coupled with the local maximums meant that simple gradient accent type methods as used by others to solve image lidar registration [2], [8], [6] could not be used here. To solve these problems particle swarm optimization was used [14], [15]. Particle swarm optimization works by placing an initial population of particles randomly in the search space. Each iteration a particle moves to a new location based on three factors: i) it moves towards the best location found by any particle, ii) it moves towards the best location it has ever found

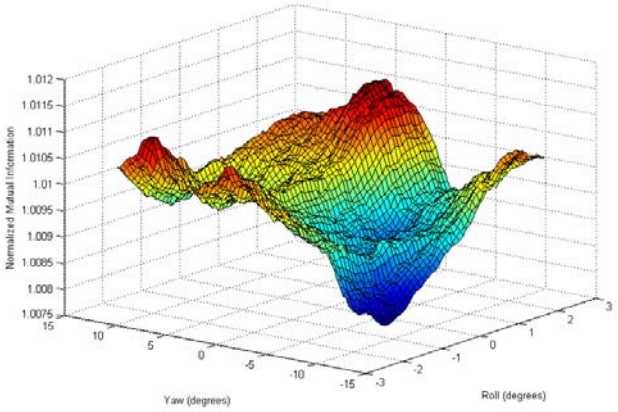

Fig. 4. Example of NMI values for changing roll and yaw

itself and iii) it moves in a random direction. The optimizer stops once all particles have converged. The entire algorithm for registration is shown in section III-E [15].

E. Algorithm

Let

$\boldsymbol{r^i(t)}$ be the position of particle i at time t

$\boldsymbol{v^i(t)}$ be the velocity of particle i at time t

$\boldsymbol{p_n^{i,L}}$ be the local best of the ith particle for the nth dimension

$\boldsymbol{p_n^g}$ be the global best for the nth dimension

$\boldsymbol{n \in 1, 2, ...N}$

$\boldsymbol{t}$ is the time

$\boldsymbol{\Delta t}$ is the time step

$\boldsymbol{c_1}$ and $\boldsymbol{c_2}$ are the cognitive and social factor constants

$\boldsymbol{\phi_1}$ and $\boldsymbol{\phi_2}$ are two statistically independent random variables uniformly distributed between 0 and 1

$\boldsymbol{w}$ is the inertial factor

for *each itertaion* l **do**
  if $f(r^i(l+1)) > f(p^{i,L}(l))$ **then**
    $p^{i,L}(l+1) = r^i$
  end
  if $f(r^i(l+1)) > f(p^g(l))$ **then**
    $p^g(l+1) = r^i$
  end
  $v_n^i(t+\Delta t) =$
  $wv_n^i(t) + c_1\phi_1[p_n^{i,L} - x_n^i(t)]\Delta t + c_2\phi_2[p_n^g - x_n^i(t)]\Delta t$
  $r_n^i(t+\Delta t) = r_n^i(t) + \Delta t v_n^i(t)$
end

IV. Results

The method was tested on three different datasets. The first dataset was obtained from an open pit mine in Western Australia containing detailed scans and hyperspectral images of cliff faces. A second dataset was obtained next to the

Australian Centre for Field robotics (ACFR) building. The third dataset used was the KITTI dataset [16] which contains velodyne scans and greyscale images from a moving car.

For all datasets the particle swarm optimizer was started with 200 particles and run until the particles all converged to within 0.1 in all dimensions of each other. This usually took 100 to 200 iterations

The code was written in Matlab with mex files written in c++ and CUDA created for the generation of the lidar images and mutual information calculations. The code was run on a dell latitude E6150 laptop with an Intel i5 M520M CPU and a NVS3100 GPU. Each function evaluation took around 0.01 seconds. The total runtime for the code was 3 to 10 minutes for the mine and ACFR dataset and 2 to 4 hours for a series of 100 image-scan pairs used in the KITTI dataset.

A. Mine experiment

The method was tested on a dataset of an open pit mine in Western Australia previously used in [17]. The feature used to colour the lidar image was the normals. This was used as no intensity of return was available. The areas where no readings were given by the lidar were coloured black and used in the registration as the strong skyline helped the alignment converge to the correct solution.

The laser used was a Riegl LMS-Z420i and the hyperspectral camera was a Neo HySpex VNIR and SWIR, the setup is shown in figure 5. RTK GPS was used to provide the exact location of the camera and laser scanner, however at two of the locations (a2 and a3) this signal failed and only a standard GPS location was given. The hyperspectral camera was readjusted and the focal length changed before taking each image so its intrinsics cannot be assumed to be the same between images. Scan and image pairs from four different sections of the mine were used. These images were taken over the course of two days. These areas of the mine were labelled a1,a2,a3,a4. An initial guess at the orientation of the camera was made. This guess was chosen such that a comparison of the hyperspectral and initial lidar scan could clearly show the alignment to be incorrect by a few degrees. The initial location of the camera was taken to be the GPS coordinates. The initial guess of the values can be found in table I

1) Optimization: Each optimization was run twice, each time with different set of parameters θ. Equation 10 shows the configurations evaluated.

$$\begin{aligned} \theta_1 &= [roll, pitch, yaw, c] \\ \theta_2 &= [roll, pitch, yaw, c, x, y, z] \end{aligned} \tag{10}$$

The search space for the optimizer was constructed based on the assumptions:

- The roll, pitch and yaw of the camera were within 10, 20 and 5 degrees respectively of the lasers.
- The cameras principal distance was within 20 pixels of correct (for this camera principal distance $\approx$ 1320).
- The x, y and z coordinates were either correct or within 4, 4 and 0.5 meters of correct.

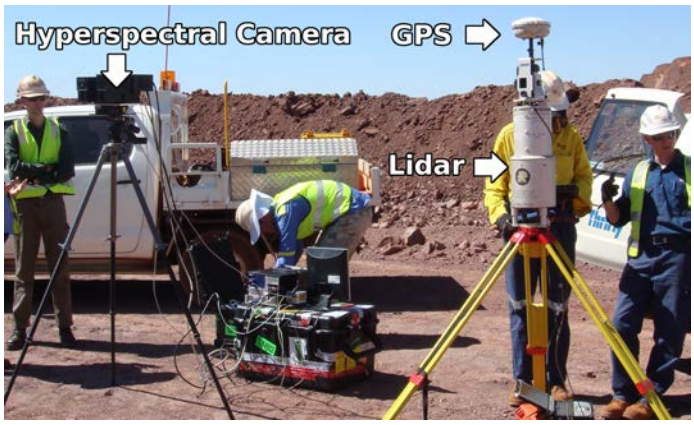

Fig. 5. Hyperspectral camera and lidar setup used to collect data

θ	site	Δx	Δy	Δz	roll	pitch	yaw	c
0	a1	0	0	0	0	0	-44	1320
1	a1	0	0	0	-0.1	0.1	-45.1	1332
2	a1	-0.138	-0.109	-0.196	-0.82	0.1	-45.2	1334
0	a2	0	0	0	0	0	37	1320
1	a2	0	0	0	-2.6	5.9	39.1	1321
2	a2	0.230	0.895	0.265	-2.5	5.6	40.4	1315
0	a3	0	0	0	0	5	5	1320
1	a3	0	0	0	-1.1	6.1	5.1	1326
2	a3	-0.02	-0.03	0.459	-1.1	6.1	5.1	1326
0	a4	0	0	0	0	4	50	1320
1	a4	0	0	0	-1.1	3.3	54.1	1337
2	a4	-0.098	-0.386	-0.319	-1.2	3.7	53.8	1338

TABLE I
PARAMETERS FOUND FOR CAMERA MODEL CAMERA MODEL. θ_0 IS THE INITIAL VALUES

2) Mine results: For the mine dataset no ground truth as to the orientation between the lidar and the hyperspectral camera was available. This means that no quantitative evidence as to the accuracy of the alignment can be given. However some indication as to the effectiveness of the method can be gained by viewing of the data[1].

The parameters that were found are listed in table I. In the absence of ground truth values few conclusions can be drawn from the table alone and so the results of four different scans are shown in figure 7. For a4 an image generated by using the calibrated camera to colour the point cloud is shown in figure 6, this gives some indication to the accuracy as the only points clearly miscoloured are caused by lidar returns off dust and tape that was blowing in the wind. Note when viewing the outputs that while a tripod with the hyperspectral calibration board (for reflectance) is present in most of the images, it was often moved between the time the hyperspectral image and the lidar scan were taken meaning it cannot be used to judge the quality of the alignment. On visual inspection it can clearly be seen that for all runs the approach converges to a solution that appears correct. For the areas a4 and a1 the RTK GPS was operating and so any variation in position can be taken as error. For dataset a1 the results are within 0.4m of each

[1]All the results are shown in movie attached

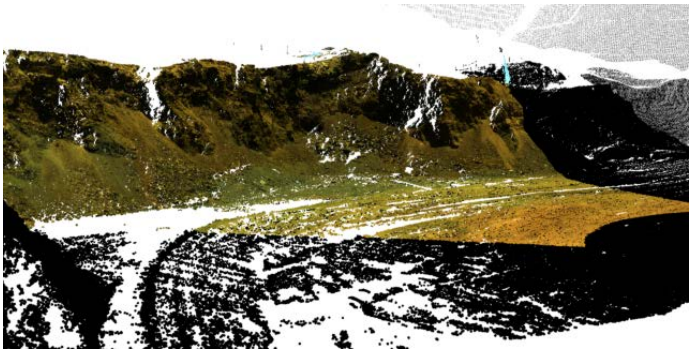

Fig. 6. Point cloud of a4 coloured by the visible bands of the calibrated hyperspectral image

Scan	roll	pitch	yaw	x	y	z	c
2	190.4	-2.4	-90.0	-0.59	0.46	0.03	763.3
3	-51.0	1.4	-91.1	-0.42	0.21	0.62	788.4
4	-30.7	-0.9	-89.6	-0.39	0.53	0.05	755.0
measured	-	-	-	≈-0.6	≈0.2	≈0.2	-

TABLE II
CALIBRATION VALUES FOR ACFR DATASET

other while a4 the error is 0.5m. These errors in translation were expected since in the mining dataset does not have close objects (25-30 metres to target) which makes the cost function less sensitive to errors in translation. This problem was also noted in [6].

B. ACFR experiment

A Specim hyperspectral camera and Riegl VZ1000 lidar scanner were mounted on top of a Toyota Hilux and used to take a series of 4 scans of the ACFR building from the park next to it, the setup is shown in figure 8. The scanner output gave the location of each point as its latitude, longitude and altitude. While the location of the scanner was recorded its orientation was not. The focal length of the hyperspectral camera was adjusted between each scan. The intensity of return were taken as features and areas with no return were not included in the NMI calculation.

The search space for the optimizer was constructed assuming the following:

- The roll, pitch and yaw of the camera were within 10, 10 and 5 degrees respectively of the lasers.
- The cameras principal distance was within 50 pixels of correct (for this camera principal distance $\approx$ 770).
- The x, y and z coordinates were within one meter of correct.

1) ACFR results: Of the four scans taken, three had solutions that converged. The last scan contained large amounts of shadow that obscured many of the features and may have been the cause of its failure. The calibration values obtained can be seen in Table II. As the orientation of the vehicle when taking the scans was not recorded no information as to the accuracy of the angle can be given. However as the sensors are in a fixed relative position if the sensors are correctly aligned the

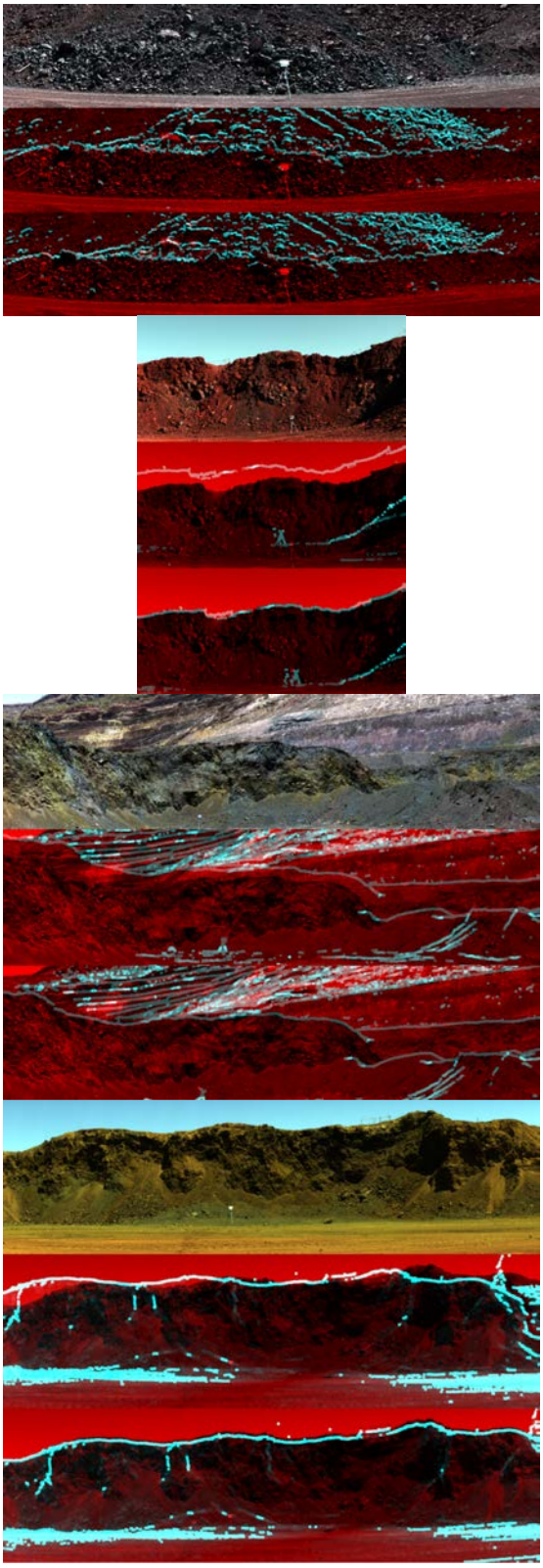

Fig. 7. Results of optimization. To visualize results strong edges in the lidar data have been found and are overlaid in blue on top of the red hyperspectral image. Each area has shown the rgb bands of the hyperspectral (top), the initial guess (middle) and the alignment found changing θ_2. The sites from top to bottom are a1, a2, a3 and a4.

Fig. 8. Setup used to collect ACFR data

Fig. 9. Calibrated camera image of ACFR projected onto point cloud

x,y,z values should all correspond to the distance between the sensors and any difference in them can be taken as error in the method.

The accuracy of the results were visually checked by projecting the calibrated camera image onto the point cloud as is shown in figure 9. Any misregistration can then be seen from the miscolouring of objects. Visually the results appear fairly accurate as even for thin objects such as light poles and tree trunks the colour is mostly correct, this goes against what table II shows though as significant differences in the location of the sensors epically in the z direction can be seen. This is due to the small effect that displacements of this magnitude have on the image. This issue is further compounded by the very similar effect that focal length and a shift in z have leading to the larger difference in z. The issue may be reduced by using a camera with known intrinsics or by taking a large series of scans and processing them all simultaneously as is done for the KITTI dataset.

C. KITTI experiment

Due to the lower resolution of the velodyne it was found that roughly 95% of the pixels in each image generated had no laser data, because of this only the areas where laser readings were given were considered in calculating the NMI. The normals were used as features, as for the intensity of return each laser gave significantly different readings for the same surface

calibration	roll	pitch	yaw	x	y	z
ground truth	-89.56	0.06	-89.87	-0.007	-0.063	-0.267
calculated	-89.72	0.70	-90.13	-0.013	-0.087	-0.314

TABLE III
CALIBRATION VALUES FOR KITTI DATASET

making this information of limited use. In the KITTI dataset the lidar was a Velodyne HDL-64E, the car was equipped with 4 cameras of these only the leftmost grayscale camera was used [16]. The dataset had a camera to lidar calibration provided that had been obtained using the method outlined in [18]. This was further refined by the manual selection of related points. A projection matrix giving the intrinsics of the camera was also provided.

The test was conducted by taking every 20th frame from the first 1000 frames of drive 71. This drive was through a very busy area with a large number of pedestrians crossing the road and passing close by. This gave the method its best chance of success as the numerous close objects ensured that an incorrect location of the camera would suffer from easily noticeable parallax error. The search space for the optimization was set so that x, y and z were within 0.5 meters of correct. Roll and yaw angles were within 15 degrees of correct and pitch angle was within 3 degrees. The small search space with pitch was to compensate for an issue encountered with the velodyne. Due to the fairly small vertical angle the scan covered for pitch angles much outside this range only a few points overlap with the image. These points tended to all have roughly the same normal values and so resulted in a high NMI value if aligned with a section of solid colour in the camera image. The parameters for the intrinsics of the camera model were taken from the KITTI calibration and assumed to be correct.

1) KITTI results: The values converged to as well as the ground truth calibration provided can be seen in table III. The method achieved a registration that was within 60mm and 1 degree of the ground truth. However, on close visual inspection of the data the alignment can be seen to in some respects be superior to that taken as the ground truth. This is demonstrated in figure 10.

V. COMPARISON WITH OTHER METHODS

An implementation of A Mastin *et al* method of using the depth image was created [2] and tested on the ACFR and mine dataset, however no images were successfully aligned using this method. This failure was expected as in most cases only a very rough guess as to the correct calibration was given where the error far exceeded the assumptions of accuracy the method made. This meant that the gradient accent optimizer failed to find the global maximum. In the velodyne dataset at the time the testing was being done, only one other method [5] was found that attempted calibration under similar conditions. This method was able to achieve accurate calibration in five minutes, significantly faster than the proposed method. It also successfully operated over a slightly larger search space. The drawback of this method however, is that it cannot be used for

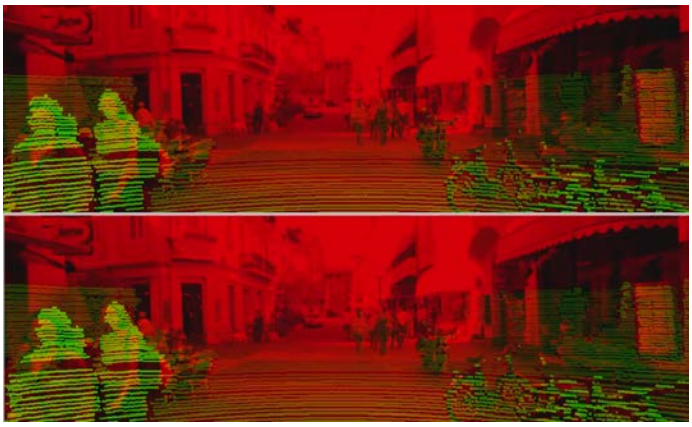

Fig. 10. A frame from the KITTI dataset with velodyne data overlaid in green. Optimized calibration on top, ground truth on bottom. Note the slight misalignment of the faces of the people on the left

a single image-scan pair. This is due to a step in the method where an average of all the images is subtracted from each image.

VI. Conclusion

A method for aligning images with lidar scans of the same area was presented. This method operates by creating an image using a camera model and coloured using either the normals or return intensities. Normalized mutual information is then used to compare the images and maximized to find the best alignment. This method is demonstrated to successfully work on three datasets with very different environments and sensors showing that it can be applied for a wide range of applications.

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] H. Li, C. Zhong, and X. Huang, "Reliable Registration of Lidar Data and Aerial Images without Orientation Parameters," *Sensor Review*, vol. 32, no. 4, 2012. [Online]. Available: http://www.emeraldinsight.com/journals.htm?articleid=17038519\&show=abstract

[2] A. Mastin, J. Kepner, and J. Fisher III, "Automatic registration of LIDAR and optical images of urban scenes," *Computer Vision and Pattern ...*, pp. 2639–2646, 2009. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5206539

[3] S. Lee, S. Jung, and R. Nevatia, "Automatic integration of facade textures into 3D building models with a projective geometry based line clustering," *Computer Graphics Forum*, vol. 21, no. 3, 2002. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1111/1467-8659.00701/full

[4] L. Liu and I. Stamos, "A systematic approach for 2D-image to 3D-range registration in urban environments," *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8, 2007. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4409215

[5] J. Levinson and S. Thrun, "Automatic Calibration of Cameras and Lasers in Arbitrary Scenes," in *International Symposium on Experimental Robotics*, 2012, pp. 1–6.

[6] 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 ...*, vol. 26, pp. 2053–2059, 2012. [Online]. Available: http://www.aaai.org/ocs/index.php/AAAI/AAAI12/paper/viewPDFInterstitial/5029/5371

[7] M. Corsini, M. Dellepiane, F. Ponchio, and R. Scopigno, "Image to Geometry Registration: a Mutual Information Method exploiting Illumination-related Geometric Properties," *Computer Graphics ...*, vol. 28, no. 7, 2009. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8659.2009.01552.x/full

[8] R. Mishra and Y. Zhang, "A Review of Optical Imagery and Airborne LiDAR Data Registration Methods," *The Open Remote Sensing Journal*, vol. 5, pp. 54–63, 2012. [Online]. Available: http://benthamscience.com/open/tormsj/articles/V005/54TORMSJ.pdf

[9] R. B. Rusu, "Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments," *KI - Künstliche Intelligenz*, vol. 24, no. 4, pp. 345–348, Aug. 2010. [Online]. Available: http://www.springerlink.com/index/10.1007/s13218-010-0059-6

[10] J. Pluim, J. Maintz, and M. Viergever, "Mutual-information-based registration of medical images: a survey," *Medical Imaging, IEEE*, vol. 22, no. 8, pp. 986–1004, 2003. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1216223

[11] C. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948. [Online]. Available: http://dl.acm.org/citation.cfm?id=584093

[12] C. Studholme, D. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern recognition*, vol. 32, no. 1, pp. 71–86, Jan. 1999. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0031320398000910http://scholar.google.com/scholar?hl=en\&btnG=Search\&q=intitle:An+overlap+invariant+entropy+measure+of+3D+medical+image+alignment\#0

[13] D. Schneider and H. Maas, "Geometric modelling and calibration of a high resolution panoramic camera," *Optical 3-D Measurement Techniques VI*, 2003. [Online]. Available: http://tu-dresden.de/die_tu_dresden/fakultaeten/fakultaet_forst_geo_und_hydrowissenschaften/fachrichtung_geowissenschaften/ipf/photogrammetrie/publikationen/pubdocs/2003_Schneider_Maas_Opt3D2003.pdf

[14] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Proceedings of ICNN'95 - International Conference on Neural Networks*, vol. 4, pp. 1942–1948, 1995. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=488968

[15] S. M. Mikki and A. a. Kishk, *Particle Swarm Optimization: A Physics-Based Approach*, Jan. 2008, vol. 3, no. 1. [Online]. Available: http://www.morganclaypool.com/doi/abs/10.2200/S00110ED1V01Y200804CEM020

[16] A. Geiger and P. Lenz, "Are we ready for autonomous driving? the kitti vision benchmark suite," *IEEE Conf. on Computer Vision ...*, pp. 3354–3361, Jun. 2012. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6248074http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6248074

[17] J. Nieto, S. Monteiro, and D. Viejo, "3D geological modelling using laser and hyperspectral data," *Geoscience and ...*, pp. 4568–4571, 2010. [Online]. Available: http://www-personal.acfr.usyd.edu.au/sildomar/files/Nieto_IGARSS_2010.pdf

[18] A. Geiger, F. Moosmann, O. Car, and B. Schuster, "Automatic camera and range sensor calibration using a single shot," *2012 IEEE International Conference on Robotics and Automation*, pp. 3936–3943, May 2012. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6224570