Direct Visual Localisation and Calibration for Road Vehicles in Changing City Environments

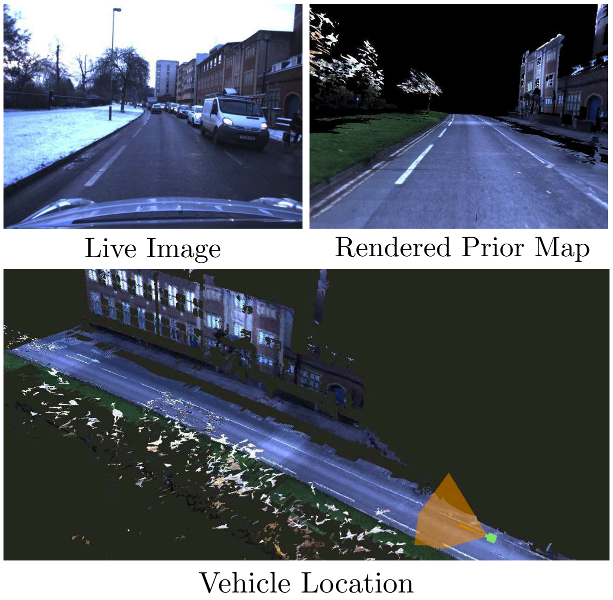
Geoffrey Pascoe
The University of Oxford
Oxford, UK
gmp@robots.ox.ac.uk

Will Maddern
The University of Oxford
Oxford, UK
wm@robots.ox.ac.uk

Paul Newman
The University of Oxford
Oxford, UK
pnewman@robots.ox.ac.uk

Abstract

This paper presents a large-scale evaluation of a visual localisation method in a challenging city environment. Our system makes use of a map built by combining data from LIDAR and cameras mounted on a survey vehicle to build a dense appearance prior of the environment. We then localise by minimising the normalised information distance (NID) between a live camera image and an image generated from our prior. The use of NID produces a localiser that is robust to significant changes in scene appearance. Furthermore, NID can be used to compare images across different modalities, allowing us to use the same system to determine the extrinsic calibration between LIDAR and camera on the survey vehicle. We evaluate our system with a large-scale experiment consisting of over 450,000 camera frames collected over 110km of driving over a period of six months, and demonstrate reliable localisation even in the presence of illumination change, snow and seasonal effects.



Live Image

Rendered Prior Map

Vehicle Location

Figure 1. Direct visual localisation with a dense appearance prior. We use a robust whole-image approach to compare the live image (top left) with a rendered 3D textured prior map (top right) collected as part of a survey. At run-time, we are able to recover the location of the vehicle (bottom) using only a single camera, despite extreme changes in illumination and weather conditions (e.g. morning snow matched to sunny afternoon). We use the same approach over multiple images to calibrate the sensors on the survey vehicle.

1. Introduction

The success of future autonomous vehicles depends on their ability to navigate from one place to another to perform useful tasks, which necessitates some form of prior map. The progress of Google's fleet of map-based autonomous vehicles [29] and the recent purchase of Nokia HERE maps by a consortium of automotive manufacturers [30] illustrate the importance of prior maps for autonomous road vehicles. A key requirement for these vehicles will be the ability to reliably localise within their prior maps regardless of the lighting and weather conditions.

To date, successful map-based autonomous vehicles have typically used 3D LIDAR sensors [15], which are robust to most outdoor lighting and weather conditions but significantly increase the cost of the sensors onboard the vehicle. Vision-based approaches have the potential to significantly reduce the sensor cost, but reliable visual localisation in the range of illumination and weather conditions experienced in a road network environment remains a challenging open problem.

In this paper we present a large-scale evaluation of a visual localisation system designed to handle changes in scene appearance due to illumination intensity, direction and spectrum as well as weather and seasonal change. We employ a 'direct' approach where all pixels from the input image are used for localisation (in contrast to interest point methods), and use a robust metric based on Normalised Information Distance to compare the input image to a dense textured prior map of the scene. We show how the pose of the camera relative to the prior map can be recovered using an optimisation framework and how incorrect localisations can be rejected using odometry information, and how the same optimisation framework can be adapted to perform camera-LIDAR calibration for a survey vehicle.

We evaluate our system in a challenging large-scale experiment using over 112km of video data from a vehicle platform in a city environment collected over a period of six months. Over this period of time we capture slow seasonal changes in the environment, along with local dynamic objects and weather conditions. We believe successful visual localisation results for experiments of this scale (both in distance and duration) demonstrate promising progress towards vision-only autonomous driving.

2. Related Work

Visual localisation in the presence of scene and illumination change is a widely studied topic; in this section we present related work focusing on visual localisation in large-scale outdoor environments towards autonomous road vehicle applications.

Successful methods for performing outdoor visual odometry, localisation and SLAM have typically made use of robust point feature detectors and descriptors, such as the well-known SIFT [17] and SURF [5]. Although effective in applications including large-scale outdoor reconstruction [1] and loop closure detection [8], these features have limitations when used over long periods of time [12], large viewpoint changes [11] and environmental change [31]. Recent attempts to learn robust descriptors [14] have led to large-scale vision-based autonomous driving demonstrations [33], but these methods require large amounts of training data from the same environment under different conditions, which can be expensive to collect.

An alternative to point-based features for long-term outdoor localisation was presented in [24], where an industrial vehicle operated autonomously in an urban environment using an architectural survey consisting of line models of buildings as a prior map. By optimising the camera exposure settings for the localisation task [23], the vehicle was able to localise reliably in extremely challenging lighting conditions using edge features alone. However, this method relies entirely on man-made structures with strong edges, which is not guaranteed for road environments.

Recent dense and semi-dense approaches to visual localisation and SLAM offer promising performance in challenging conditions. The method in [21] builds a dense scene representation using all pixels from the input camera images, and demonstrates robustness to significant viewpoint changes and motion blur, but has only been tested in small indoor environments. Semi-dense approaches such as [9] and [10] have been tested in larger outdoor environments, but only across short periods of time. Both these methods use a photometric cost function that directly compares pixel intensities between the camera images and the map, and are therefore not robust to changes in lighting direction, intensity and spectrum that occur in typical outdoor environments over time.

To perform direct (whole-image) localisation in the presence of illumination changes, many researchers have made use of information-theoretic methods. Originally adopted by the medical imaging community to align images from different modalities [19], information-theoretic methods have become popular for matching camera images under different conditions. Caron *et al.* [6] performed mutual-information-based image alignment using a vehicle-mounted camera and a commercial 3D model of a city, evaluated on a number of short trajectories. Wolcott *et al.* [32] demonstrated $\mathbb{SE}(2)$ visual localisation on an autonomous vehicle platform using a LIDAR-based appearance prior. Localisation in $\mathbb{SE}(3)$ was demonstrated using a 3D point-cloud in [28] and using a textured mesh in [27], however both of these were only evaluated on short trajectories in urban environments.

Methods used to localise cameras across different modalities have also enabled targetless camera-LIDAR calibration approaches. Napier *et al.* [20] generate synthetic LIDAR reflectance images and align gradients with camera images; the calibration is updated over time using a recursive Bayesian process. Pandey *et al.* [25] perform an optimisation process to maximise the mutual information between the camera image and reprojected LIDAR points. These methods offer promising results towards lifelong visual localisation and calibration; in the following sections we present our progress towards these goals.

3. Normalised Information Distance

In this section we briefly summarise an approach to direct visual localisation using the Normalised Information Distance (NID) metric [16]. Given a camera image I_C, we seek to recover the most likely $\mathbb{SE}(3)$ camera location G_{CW} in the world frame W by minimising the following objective function:

$$\hat{G}_{CW} = \underset{G_{CW}}{\arg\min}\, \text{NID}\left(I_C, I_S\left(G_{CW}, S_W\right)\right) \tag{1}$$

where S_W is the world-frame scene geometry and texture information, dubbed the *scene prior*, and I_S is the rendering function that produces a synthetic image from location G_{CW}. An example camera image I_C and corresponding synthetic image I_S is shown in Fig. 1. The NID metric is defined as follows:

$$\mathrm{NID}\left(I_C, I_S\right) = \frac{2\mathrm{H}\left(I_C, I_S\right) - \mathrm{H}\left(I_C\right) - \mathrm{H}\left(I_S\right)}{\mathrm{H}\left(I_C, I_S\right)} \tag{2}$$

The terms $\mathrm{H}\left(I_C\right)$ and $\mathrm{H}\left(I_S\right)$ represent the marginal entropies of I_C and I_S and $\mathrm{H}\left(I_C, I_S\right)$ is the joint entropy of I_C and I_S, defined as follows:

$$\mathrm{H}\left(I_C\right) = -\sum_{a=1}^{n} p_C\left(a\right)\log\left(p_C\left(a\right)\right) \tag{3}$$

$$\mathrm{H}\left(I_C, I_S\right) = -\sum_{a=1}^{n}\sum_{b=1}^{n} p_{C,S}\left(a,b\right)\log\left(p_{C,S}\left(a,b\right)\right) \tag{4}$$

where p_C and $p_{C,S}$ are the marginal and joint distributions of the images I_C and I_S, represented by n-bin discrete histograms where a and b are individual bin indices. $\mathrm{H}\left(I_S\right)$ is defined similarly to (3) for I_S. Further derivation of the discrete distributions and their derivatives can be found in [26].

The key advantage of the NID metric over direct photometric approaches is that the distance is not a function of the pixel *values* in the camera and reference images, but a function of the *distribution* of values in the images [16]. Hence, NID is robust not only to global image scaling (due to camera exposure changes) but also arbitrary colourspace perturbations. As will become evident, this property makes NID particularly suited to matching images under different conditions, including changes in lighting intensity, direction and spectrum.

The normalisation property of NID offers similar advantages over conventional mutual information (MI) metrics, as it does not depend on the total information content of the two images; hence it will not favour matches between highly textured image regions to the detriment of the global image alignment [28].

4. Scene Prior Construction

To produce the world-frame scene geometry and texture information S_W required for localisation, we employ a mapping process combining cameras, low-cost 2D LIDAR and odometry information. We assume that a subset of 'survey' vehicles equipped with such sensors traverse the road network infrequently (similar to Google Street View [3]), providing all road users with a 3D prior map that reflects the structure and appearance of the environment at survey time. Fig. 2 illustrates the sensor configuration of the survey vehicle.

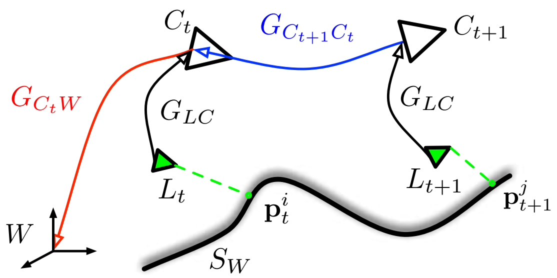


Figure 2. Vehicle configuration for mesh generation. The vehicle is equipped with a camera C and LIDAR L related by the calibration transform G_{LC}. As the LIDAR moves through the world it observes points $\mathbf{p}_t^i$; the collection of all point geometry and texture information makes up the scene prior S_W. The mapping process involves generating the scene prior S_W using camera images I_{C_t} and LIDAR scans L_t using known world frame locations G_{C_tW}; the localisation process involves recovering the camera locations G_{C_tW} using only camera images I_{C_t} and the scene prior S_W.

An additional consideration is the time synchronisation of all sensors on the survey vehicle; this is critical to ensure points measured from the LIDAR reproject to the correct location in the camera when the vehicle is in motion. We employ the TICSync library [13] to guarantee accurate synchronisation between camera and LIDAR timestamps.

4.1. Scene Prior Geometry

The geometry of the scene S_W is represented by a triangle mesh, chosen for ease of use in graphics and rendering pipelines. The 2D LIDAR is mounted to the vehicle in a vertical 'push-broom' configuration [4], and therefore experiences significant out-of-plane motion. Each LIDAR scan returns points $\mathbf{p}_t^i$ at index i in the local frame L_t; these points can be transformed to the world frame ${}^W\mathbf{p}_t^i$ as follows:

$${}^W\mathbf{p}_t^i = G_{C_tW}G_{LC}\mathbf{p}_t^i \tag{5}$$

Due to the prismatic nature of push-broom LIDAR scans, a triangle mesh can be formed by stitching together successive scans based on point index. We accept triangles ${}^W\mathbf{t} = \{{}^W\mathbf{p}_1, {}^W\mathbf{p}_2, {}^W\mathbf{p}_3\}$ from triplets of points that satisfy the following conditions:

$$\max \left\| \begin{matrix} {}^W\mathbf{p}_t^i - {}^W\mathbf{p}_t^{i+1} \\ {}^W\mathbf{p}_t^{i+1} - {}^W\mathbf{p}_{t+1}^i \\ {}^W\mathbf{p}_t^i - {}^W\mathbf{p}_{t+1}^i \end{matrix} \right\| < \delta, \text{or} \tag{6}$$

$$\max \left\| \begin{matrix} {}^W\mathbf{p}_t^{i+1} - {}^W\mathbf{p}_{t+1}^i \\ {}^W\mathbf{p}_t^{i+1} - {}^W\mathbf{p}_{t+1}^{i+1} \\ {}^W\mathbf{p}_{t+1}^i - {}^W\mathbf{p}_{t+1}^{i+1} \end{matrix} \right\| < \delta \tag{7}$$

The parameter δ represents the maximum acceptable triangle edge length; for urban environments we found $\delta = 1$m to be an acceptable threshold for meshing without interpolating triangles between different objects. Fig. 3(b) illustrates a typical mesh generated using this process.

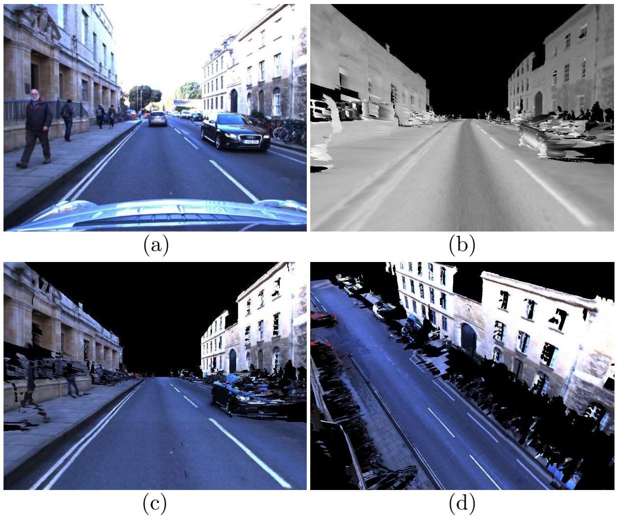


Figure 3. Stages of the map building process. (a) Camera image for a single position in the map dataset. (b) Geometric information for map obtained from LIDAR data. (c) Mesh textured with camera images. (d) Novel viewpoint synthesised from textured mesh.

4.2. Scene Prior Appearance

To capture the scene appearance information at survey time, we texture the mesh using camera images collected as part of the survey process. We use a projective texturing approach, where for each camera image I_{C_t} we project every triangle ${}^{W}\mathbf{t}$ onto the camera frame:

$$ {}^{C_t}\mathbf{t} = K G_{C_t W}^{-1} \, {}^{W}\mathbf{t} \tag{8} $$

where K is the perspective projection matrix of the camera. For each triangle ${}^{W}\mathbf{t}$ we find the camera pose C_t that maximises the area (in pixels) of the reprojected triangle ${}^{C_t}\mathbf{t}$; we then store the corresponding portion of I_{C_t} as the texture information for ${}^{W}\mathbf{t}$. Using this textured mesh as the scene prior S_W we can generate synthetic images $I_S\left(G_{CW}, S_W\right)$ at any location G_{CW} as required for (1). Fig. 3(d) illustrates a textured mesh generated using this process.

5. Monocular Localisation

At run-time we wish to localise a vehicle within the scene S_W using a single monocular camera C. Starting from an initial guess $\hat{G}_{CW}$ provided by the previous vehicle position or a weak localiser such as GPS or image retrieval [7], we seek to solve (1) using an optimisation framework. Concretely, we iteratively solve the following quasi-Newton objective:

$$ \mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k B_k^{-1} \frac{\partial \text{NID}\left(I_C, I_S\left(g\left(\mathbf{x}_k\right), S_W\right)\right)}{\partial \mathbf{x}_k} \tag{9} $$

where $\mathbf{x}$ is a minimal 6-parameterisation of the $\mathbb{SE}(3)$ camera location G_{CW} and the function $g\left(\mathbf{x}\right) = G_{CW}$. B_k represents the Hessian approximation at time k, obtained using the BFGS method [22], and α_k is the step distance resulting from a robust line search. We compute analytical derivatives of the NID cost function using the method in [26] and solve using the BFGS implementation in Ceres Solver [2].

5.1. Outlier Rejection

Due to dynamic objects occluding the scene or extreme changes in appearance, the optimisation in (9) will occasionally either fail to converge or converge to an incorrect location. To increase robustness we use the odometry information $G_{C_{t+1}C_t}$ to verify the consistency between successive localisations G_{C_t} and $G_{C_{t+1}}$. We first use the odometry to estimate the camera location $\hat{G}_{C_{t+1}}$ as follows:

$$ \hat{G}_{C_{t+1}W} = G_{C_t W} G_{C_{t+1}C_t} \tag{10} $$

We then make use of the final Hessian approximation B_k from (9) at time t, denoted B_t, to estimate the uncertainty of $\hat{G}_{C_t W}$ and propogate it to time $t+1$:

$$ \hat{\Sigma}_{t+1} = \frac{\partial \hat{G}_{C_{t+1}W}}{\partial \mathbf{x}_t} B_t^{-1} \frac{\partial \hat{G}_{C_{t+1}W}}{\partial \mathbf{x}_t}^{\mathrm{T}} \tag{11} $$

where $\hat{\Sigma}_{t+1}$ is the estimated covariance matrix at time $t+1$. Note that we assume the uncertainty of $G_{C_{t+1}C_t}$ is small compared to the localisation uncertainty. From this we can compute the squared Mahalanobis distance s between the localisation estimate predicted using odometry $\hat{G}_{C_t W}$ and the one obtained by camera localisation $G_{C_t W}$ as follows:

$$ \mathbf{e} = g^{-1}\left(\hat{G}_{C_{t+1}W}^{-1} G_{C_{t+1}W}\right) \tag{12} $$

$$ s = \mathbf{e}^{\mathrm{T}} \left(\hat{\Sigma}_{t+1} + B_{t+1}^{-1}\right)^{-1} \mathbf{e} \tag{13} $$

If the distance s exceeds a threshold ϕ, the localisation estimate $G_{C_{t+1}W}$ is deemed to be an outlier and the odometry-only estimate $\hat{G}_{C_{t+1}W}$ is substituted. In practice we find the optimisation (9) either yields correct localisation estimates, or confidently asserts the solution is far away from the correct location. By rejecting inconsistent sequential localisation estimates as outliers rather than attempting to fuse estimates based on their Hessian approximations (which are typically overconfident), we significantly increase the robustness of the localisation process.

6. Multisensor Calibration

Both the meshing and texturing pipelines in Section 4 rely on accurate knowledge of the camera-LIDAR transform G_{LC}. However, if the survey vehicles traverse the environment frequently over many years, it is unlikely that

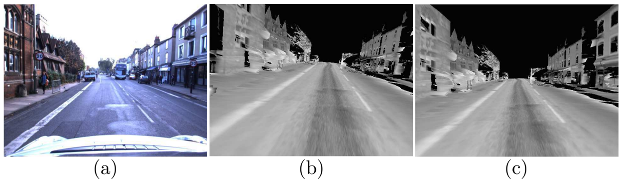


Figure 4. Camera-LIDAR calibration rendered using LIDAR reflectance information. (a) shows the reference camera image, and (b) the resulting scene prior with a calibration estimate with an orientation error less than 10 degrees. (c) shows the correct calibration recovered by the NID minimisation over a series of ten images.

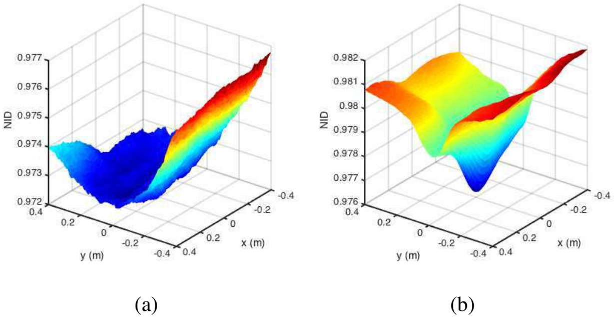


Figure 5. NID cost surface for the calibration transform G_{LC} in x and y axes for (a) one image and (b) ten images. With a single image the cost surface is rough and contains multiple minima, depending on the local structure. By optimising over multiple locations simultaneously, the cost surface is smoother and contains a distinct single minima. The effect is similar for the other four axes.

a single calibration procedure performed at the factory will be valid for the lifetime of the vehicle. A calibration error of only a few degrees can significantly degrade the scene prior geometry, as illustrated in Fig. 4. Instead, we wish to determine the optimal calibration for any survey as a post-processing stage using only data collected during normal driving.

The key property of NID enabling camera-LIDAR calibration is the ability to match between sensor modalities; in this case, the visible light wavelengths observed by the camera and the near-infrared (NIR) surface reflectance observed by the LIDAR scanner. By observing that the scene geometry S_W depends on the camera-LIDAR calibration transform G_{LC} during meshing in (5), we can pose the calibration problem as follows:

$$\hat{G}_{LC} = \underset{G_{LC}}{\arg\min} \sum_{i}^{N} \mathrm{NID}\left(I_{C_i}, I_R\left(G_{C_iW}, S_W\left(G_{LC}\right)\right)\right) \tag{14}$$

Here I_R is the rendering process that produces a synthetic image textured using LIDAR reflectance information, as illustrated in Fig. 4. By performing the optimisation across N images in different locations, we reduce the effect of individual scene geometry on the resulting calibration transform. Empirically, we have observed that $N = 10$ or more evenly spaced images (without overlap) produce smooth cost surfaces with clear minima, as shown in Fig. 5.

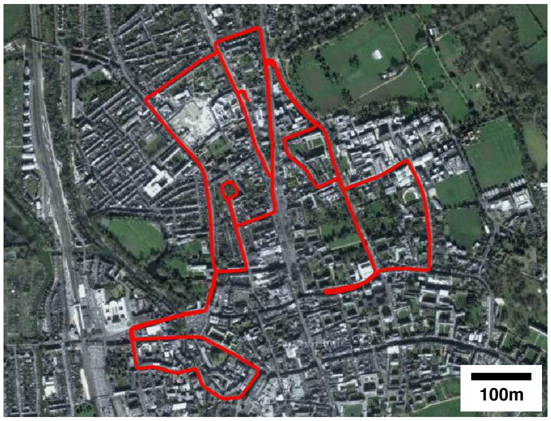


Figure 6. Dataset collection route. The route consists of twelve traversals of a 9.3km route through a city centre over a period of six months, a total of 112km of driving.

7. Evaluation

We evaluated the localisation with a large-scale experiment intended to test the system in a challenging autonomous road vehicle scenario. We collected camera, LIDAR and odometry data from an autonomous vehicle platform for twelve traversals of a 9.3km route in a city environment over a period of six months. The dataset consists of over 450,000 camera images over a total distance of 112km. The dataset collection route is shown in Fig. 6. An experiment of this duration revealed conditions not often encountered by existing visual localisation systems; seasonal effects including snowfalls and changes in deciduous trees were captured as shown in Fig. 7.

For camera-LIDAR calibration and scene prior generation, we used Dataset 1 collected in November 2014. For all remaining datasets we localised the camera relative to the scene prior for Dataset 1 using only the camera and odometry information. The traversal of Dataset 12 was performed in April 2015, 161 days after Dataset 1.

Ground truth for a dataset of this scale was challenging to obtain; although the vehicle platform was equipped with a GPS-aided INS system, the inherent drift in the GPS constellation over this period introduced significant global translational offsets between datasets. We instead performed a large-scale pose graph optimisation using all twelve datasets with loop closures provided by FAB-MAP [7] and manually corrected, yielding a globally consistent

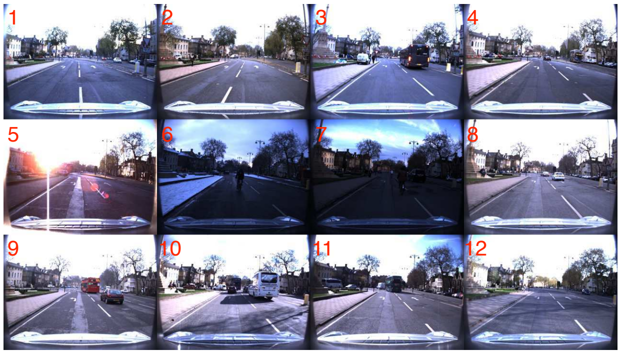


Figure 7. Camera images from a sample location for all twelve evaluation datasets. The traversals include significant variation in lighting and weather, including direct sunlight in dataset 5, snow in dataset 6 and seasonal vegetation change in dataset 9. Dataset 1 is used to build the scene prior; all subsequent datasets are matched to dataset 1.

map for offline evaluation.

The scene prior rendering, NID value and derivative calculations were all implemented in OpenGL and computed using a single AMD Radeon R295x2 GPU. Localisation on 640x480 resolution images was performed at approximately 6Hz, where each localisation consisted of approximately 20 cost function and derivative evaluations.

8. Results

In this section we present the results obtained from evaluating our localiser across the datasets shown in Fig. 7.

8.1. Localisation Performance

The localisation performance of the system is shown in Table 1. All datasets except for the one including snow (dataset 6) achieved RMS errors of less than 30cm in translation, and most had errors of less than 1° in rotation.

Fig. 8 presents histograms of the absolute translational and rotational localisation errors across all the evaluation datasets. The reduction of errors beyond 5m and 5 degrees indicates the outlier rejection method is successfully removing incorrect localisation estimates that would otherwise introduce large errors. Fig. 9 illustrates the distribution of outliers by analysing the likelihood of travelling a certain distance without a successful localisation relative to the prior. Despite an outlier rate of close to 10% on some datasets, the likelihood of travelling more than 10m without a successful localisation is less than 3%, and the worst-case distance travelled using only odometry information is 40m over the entire 100+km experiment.

8.2. Failure Cases

As shown in Table 1, close to 10% of our localisations on some datasets result in outliers. Some typical examples

Dataset	RMS Translation Error (m)	RMS Rotation Error (°)	Outlier Rate (%)
2	0.2300	0.7391	3.46
3	0.2219	0.9912	2.76
4	0.2284	1.0600	8.06
5	0.1779	0.8422	2.76
6	0.3307	1.4037	7.44
7	0.2295	1.3751	5.86
8	0.2059	0.7620	1.72
9	0.2496	0.8301	9.58
10	0.2872	0.8480	4.82
11	0.2510	0.9568	5.79
12	0.2082	1.0451	4.06

Table 1. RMS errors in translation and rotation, for each dataset. Also shown is the percentage of rejected outliers, determined as described in Section 5.1. We achieve RMS errors less than 30cm and 1.5° on all datasets except number 6, in which the scene was covered in snow.

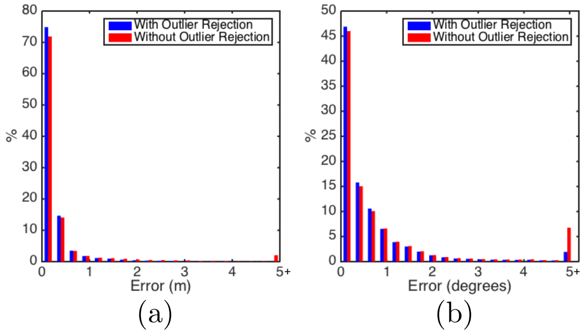


Figure 8. Error histograms for translation (a) and rotation (b) for all datasets combined. 94.5% of translational errors were under one metre in magnitude, and 79% of rotational errors were under one degree. Outlier rejection successfully removed the vast majority of errors over 5m or 5°.

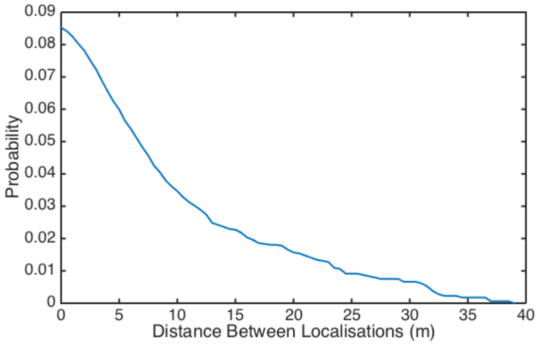


Figure 9. Probability of travelling more than a given distance without a successful localisation. The likelihood of travelling more than 10m without localisations (using odometry alone) is less than 3%.

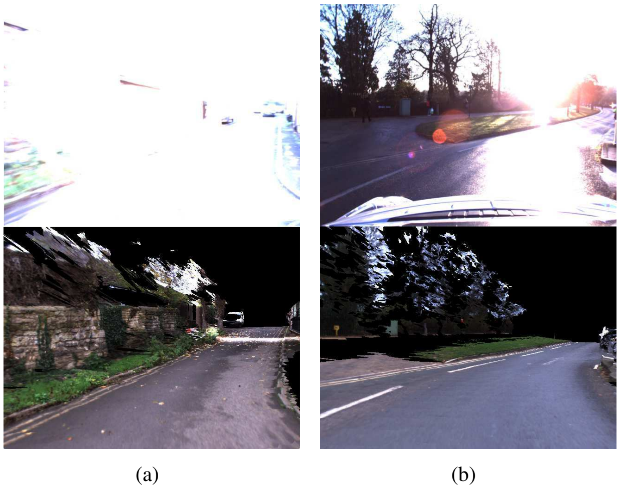


Figure 10. Failed localisation due to limited dynamic range. (a) shows a location where the camera autoexposure algorithm failed to provide a well-exposed image, and (b) shows a location with direct sunlight in the field of view. These cases could be resolved with a logarithmic or HDR camera.

of these failures are shown in Figs. 10–11.

Many of the localisation failures are due to poor quality imagery, for example over-exposure and direct sunlight as shown in Fig. 10. Imagery such as this preserves little of the structure of the scene, resulting in shallow basins in the NID cost surface.

Large occlusions, either in the map or the live camera imagery, are also a common cause of localisation failure. For example, Fig. 11(a) shows a case with a large occlusion taking up the majority of the camera image, hence causing localisation to fail. Fig. 11(b) has occlusions in both the live camera imagery and the prior, and there is little in common between the appearance of the two scenes.

9. Conclusions

We have presented a large-scale evaluation of a visual localisation approach based on Normalised Information Distance. We demonstrated its robustness and reliability in a range of challenging conditions, including snow and seasonal changes, where it was able to maintain localisation estimates within 30cm and 1.5 degrees of the true location. Additionally, we have illustrated progress towards lifelong calibration for vehicles equipped with both camera and LIDAR sensors. We hope these results highlight the challenges of visual localisation in changing environments, and pave the way for low-cost and long-term autonomy for future road vehicles.

9.1. Future Work

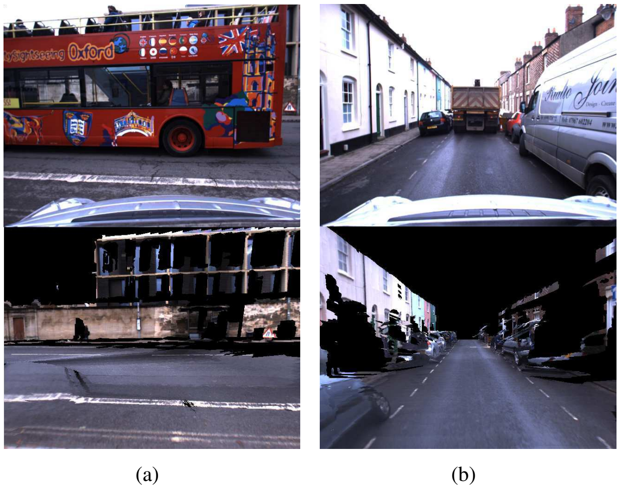


Figure 11. Failed localisation due to occlusions. (a) shows the field of view almost entirely occluded by a bus, and (b) shows a narrow street with many occluding vehicles and differences in parked cars. These cases could be resolved with a higher-level classification layer to detect and ignore dynamic objects for the purposes of localisation.

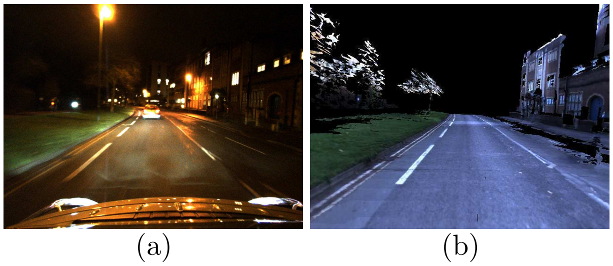


Figure 12. Failed localisation between a night-time dataset (a) and daytime prior (b). The localisation failed to converge for all images collected at night, due to the highly non-uniform illumination. A possible solution is a separate appearance prior collected at night, but this doubles storage requirements.

There remain a number of scenarios for which even NID is not able to localise; principally, localisation at night against a prior collected during the day or vica versa. We experimented with a night-time dataset and found the optimisation in (9) failed to converge for all locations, as illustrated in Fig. 12. Results in [18] present promising results towards 24-hour localisation, but required an appearance prior collected at night as well as one collected during the day and therefore doubles the storage requirements for mapping any given location. We aim to continue to test the system in more challenging scenarios (e.g. fog, heavy rain) and characterise the performance for larger datasets over longer periods of time.

References

[1] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski. Building rome in a day. *Com-*

munications of the ACM, 54(10):105–112, 2011. 2

[2] S. Agarwal, K. Mierle, and Others. Ceres Solver. 4

[3] D. Anguelov, C. Dulong, D. Filip, C. Frueh, S. Lafon, R. Lyon, A. Ogale, L. Vincent, and J. Weaver. Google street view: Capturing the world at street level. *Computer*, (6):32–38, 2010. 3

[4] I. Baldwin and P. Newman. Road vehicle localization with 2d push-broom lidar and 3d priors. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 2611–2617. IEEE, 2012. 3

[5] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *Computer vision–ECCV 2006*, pages 404–417. Springer, 2006. 2

[6] G. Caron, A. Dame, and E. Marchand. Direct model based visual tracking and pose estimation using mutual information. *Image and Vision Computing*, 32(1):54–63, 2014. 2

[7] M. Cummins and P. Newman. FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance. *The International Journal of Robotics Research*, 27(6):647–665, 2008. 4, 5

[8] M. Cummins and P. Newman. Appearance-only slam at large scale with fab-map 2.0. *The International Journal of Robotics Research*, 30(9):1100–1123, 2011. 2

[9] J. Engel, T. Schöps, and D. Cremers. Lsd-slam: Large-scale direct monocular slam. In *Computer Vision–ECCV 2014*, pages 834–849. Springer, 2014. 2

[10] C. Forster, M. Pizzoli, and D. Scaramuzza. Svo: Fast semi-direct monocular visual odometry. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pages 15–22. IEEE, 2014. 2

[11] P. Furgale and T. D. Barfoot. Visual teach and repeat for long-range rover autonomy. *Journal of Field Robotics*, 27(5):534–560, 2010. 2

[12] A. J. Glover, W. P. Maddern, M. J. Milford, and G. F. Wyeth. Fab-map+ ratslam: appearance-based slam for multiple times of day. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 3507–3512. IEEE, 2010. 2

[13] A. Harrison and P. Newman. Ticsync: Knowing when things happened. In *Proc. IEEE International Conference on Robotics and Automation (ICRA2011)*, Shanghai, China, May 2011. 05. 3

[14] H. Lategahn, J. Beck, B. Kitt, and C. Stiller. How to learn an illumination robust image feature for place recognition. In *Intelligent Vehicles Symposium (IV), 2013 IEEE*, pages 285–291. IEEE, 2013. 2

[15] J. Levinson and S. Thrun. Robust vehicle localization in urban environments using probabilistic maps. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 4372–4378. IEEE, 2010. 1

[16] M. Li, X. Chen, X. Li, B. Ma, and P. M. B. Vitanyi. The similarity metric. *Information Theory, IEEE Transactions on*, 50(12):3250–3264, 2004. 2, 3

[17] D. G. Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999. 2

[18] W. Maddern, A. D. Stewart, and P. Newman. Laps-ii: 6-dof day and night visual localisation with prior 3d structure for autonomous road vehicles. In *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, pages 330–337. IEEE, 2014. 7

[19] F. Maes, D. Vandermeulen, and P. Suetens. Medical image registration using mutual information. *Proceedings of the IEEE*, 91(10):1699–1722, 2003. 2

[20] A. Napier, P. Corke, and P. Newman. Cross-calibration of push-broom 2d lidars and cameras in natural scenes. In *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, Karlsruhe, Germany, May 2013. 2

[21] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison. Dtam: Dense tracking and mapping in real-time. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2320–2327. IEEE, 2011. 2

[22] J. Nocedal and S. J. Wright. *Numerical Optimization*, volume 43. 1999. 4

[23] S. Nuske, J. Roberts, and G. Wyeth. Extending the dynamic range of robotic vision. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 162–167. IEEE, 2006. 2

[24] S. Nuske, J. Roberts, and G. Wyeth. Robust outdoor visual localization using a three-dimensional-edge map. *Journal of Field Robotics*, 26(9):728–756, 2009. 2

[25] G. Pandey, J. R. McBride, S. Savarese, and R. M. Eustice. Automatic extrinsic calibration of vision and lidar by maximizing mutual information. *Journal of Field Robotics*, 2014. 2

[26] G. Pascoe, W. Maddern, and P. Newman. Robust Direct Visual Localisation using Normalised Information Distance. In *British Machine Vision Conference (BMVC)*, Swansea, Wales, 2015. 3, 4

[27] G. Pascoe, W. Maddern, A. D. Stewart, and P. Newman. FARLAP: Fast Robust Localisation using Appearance Priors. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Seattle, WA, USA, May 2015. 2

[28] A. D. Stewart and P. Newman. Laps-localisation using appearance of prior structure: 6-dof monocular camera localisation using prior pointclouds. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 2625–2632. IEEE, 2012. 2, 3

[29] S. Thrun. What we're driving at, 2010. 1

[30] R. Trenholm. Nokia sells Here maps business to carmakers Audi, BMW and Daimler, 2015. 1

[31] C. Valgren and A. J. Lilienthal. Sift, surf and seasons: Long-term outdoor localization using local features. In *EMCR*, 2007. 2

[32] R. W. Wolcott and R. M. Eustice. Visual localization within lidar maps for automated urban driving. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pages 176–183. IEEE, 2014. 2

[33] J. Ziegler, H. Lategahn, M. Schreiber, C. G. Keller, C. Knoppel, J. Hipp, M. Haueis, and C. Stiller. Video based localization for bertha. In *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, pages 1231–1238. IEEE, 2014. 2