

Road DNA Based Localization for Autonomous Vehicles

Liang Li, Ming Yang, Chunxiang Wang and Bing Wang

Abstract—High-precision and reliable localization is current research focus in the area of autonomous vehicles. Previous studies rely on either high-cost sensors or some specific characteristics, which means that the methods are limited to only a bit given situations. In this paper, a road DNA based localization method is proposed. It could afford high-precision result and does not have the shortcomings of previous methods at the same time. The scenery on both sides of the roads are used to generate the prior-map. The map is presented as grid map by the joint probability of occupation and reflectivity. With this type of map, different environments show different properties, which means that this method is not limited to specific environments and is effective in most cases. It costs much less memory than the previous maps. The map and live road scene flattening are both generated by data collected by low-cost LIDAR. Normalized Information Distance is utilized to align the live road scene flattening with the road DNA. Experiments show the validation and precision of this method.

I. INTRODUCTION

Autonomous vehicle has become a research hotspot in the last decades. Recently, high-precision localization for intelligent vehicles has been stressed a lot. High-precision localization complemented with some detection methods (*e.g.* pedestrian detection, obstacle detection, traffic sign detection, *etc.*) can make the vehicle intelligent or even unmanned. Therefore, it is the core technology for autonomous vehicles. Early studies about this problem can trace back to 1990s. Simultaneous Localization And Mapping (SLAM) [1] is one of the state-of-the-art methods in the early days. With good sensors, this method seem promising. While SLAM can only afford relative positioning and more emphasizes on localization in unknown environment. In known environment, localization and map correction rely on loop closure detection. Based on SLAM, some localization methods designed for intelligent vehicles have attracted increasing interest both in research and application. Due to outdoor, complicated and on-road environment of vehicles, the localization methods vary from original SLAM accordingly.

For the use of autonomous vehicles, the mapping and localization process are always separated. A prior-map is generated with the data collected by sensors first. When traveling in the area that the map covers, localization can be achieved

by matching the live observation with the prior-map. In [2], the prior-map (experience) and local swathe are produced by projecting the 2D laser scans along the vehicle trajectory. A series of samples sampled from height and reflectance in different locations are matched with the experience. The final transformation is computed by optimization method. Ryan W. Wolcott *et al.* [3] proposed the multiresolution Gaussian mixture maps which is a kind of occupied grid map in essential. These maps are a collection of Gaussian mixtures over the z-height distribution. In consideration for real time use, coarse to fine framework is utilized and the maps are divided into different resolutions. They built these maps offline as a full SLAM problem. Online localization is then performed by multiresolution search. Chong *et al.* [4] proposed synthetic 2D LIDAR to solve the localization problem on a virtual 2D plane. Features that are perpendicular to the ground are extracted by point classification. Then the prior-map is constructed using these features. A Monte Carlo Localization scheme is adopted for vehicle position estimation, based on synthetic LIDAR measurements and odometry information. An alternative method is high-precision localization based on vision-only solutions. These methods are based on the idea of visual SLAM and the cost of sensors decreases largely. The discussion of vision based localization methods is beyond this paper, if interested, please refer to [5], [6], *etc.*

From the previous studies, we can conclude that there are still many challenges in the area of high-precision localization for autonomous vehicle. The bottleneck that are encountered now are mainly:

- Size of map. The size of map, especially that built by LIDAR is too big for large scale use. Usually the size require more than 500MB per kilometer.
- Cost of the whole system. This is the limit of LIDAR based methods. The cost of 32-beam or 64-beam are too high for civil use.
- A robust point set registration method (or similarity function).

The main contribution of this paper is that a novel framework is proposed that solves the three problems above in a large part. The remainder of this paper is organized as follows. Section II presents the overview of the whole system. In Section III, the process of road DNA generation is presented. Section IV describes online localization. Experiment setup and results analysis are shown in Section V. Finally, the paper is concluded in Section VI.

II. SYSTEM OVERVIEW

This system consists of two parts. The first part is offline mapping and the second part is online localization. These two

This work was supported by the General Program of National Natural Science Foundation of China (61174178), the Major Research Plan of National Natural Science Foundation of China (91420101) and National Magnetic Confinement Fusion Science Program (2012GB102002).

Liang Li, Ming Yang and Bing Wang are with the Department of Automation, Shanghai Jiao Tong University, Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai, 200240, CN (phone: +86-21-34204533; email: MingYang@sjtu.edu.cn).

Chunxiang Wang is with Research Institute of Robotics, Shanghai Jiao Tong University, Shanghai 200240, China.

to the road surface, the road surface will not provide any useful information though the reflectance information of it maybe useful in other cases (the proposed method is mainly aimed at urban environment). The method for cutting off road surface here is Random Sample Consensus (RANSAC). The basic idea is to find a plane that can contain most points in the current cloud. A detail of the algorithm can be seen in [9]. By the RANSAC, four parameters $\{A, B, C, D\}$ of the road surface can be gotten, and the equation of that plane is $Ax + By + Cz + D = 0$. The point cloud after removing road surface can be seen in Fig.2(b). In this paper, we seek to implement an algorithm that be able to perform well regardless of environment changes. In previous studies, the relevant methods are mainly designed for structural environments which limits its wide use for real application. The biggest problem of these methods is that they did not analyze the structure of the environment, and matching the live data against the prior maps directly. In this paper, we propose road scene map which will give the important but easily ignored parts a bigger weight. For use in big areas, the road scene map is compressed in two planes on both sides of the road. After compression, the size of map is much smaller than the raw point cloud and also much smaller than most present forms of maps, which offer possibility for big area use.

First, similar to OctoMap [10], the raw point cloud is converted to 3D occupied grid map. The value in every voxel is defined by co-possibility of the nearest points in this voxel to the center of the road and the possibility of this voxel is occupied. By doing this, the lateral changes on both sides of the roads are modeled and this model is unique at a given place, which means that the kidnapping problem can be averted. It represents the world in discrete form. Every voxel store a value as is shown by equation (5):

$$\mathbf{P}(n|\mathbf{Q}) = [1 + \frac{f(n)}{N}]^{-1} \quad (5)$$

In equation (5), $\mathbf{P}(n|\mathbf{Q})$ is the value that voxel n stores given the point cloud \mathbf{Q} . $f(n)$ measures the distance between the gravity centroid of all the points in voxel n and original point in y-dimension in the vehicle coordinate frame. N is the total number of points in voxel n . So $\mathbf{P}(n|\mathbf{Q})$ is $\propto N$ and $\propto \frac{1}{f(n)}$. The 3D occupied grid map of point cloud shown in Fig.2 can be seen in Fig.3. And multi-resolution versions are also utilized here for coarse to fine localization in the next procedure, which will make the convergence faster to achieve the goal of real-time application.

But the size of this type of map is also too big for large scale use. It needs more than 10MB per kilometer, so compression is needed here. Due to this, the occupied grids are projected onto two planes which are perpendicular to road surface detected beforehand and parallel to the vehicle trajectory. The parameter of the planes are defined by:

$$AA' + BB' + CC' = 0 \quad (6)$$

$$A'A_T + B'B_T + C'C_T = 0 \quad (7)$$

$$d(P', V) = \min f(n) \quad (8)$$

In equation (6) $\{A, B, C\}$ are parameters of the road surface, and $\{A', B', C'\}$ are parameters of the plane we want to get. In equation (7), $\{A_T, B_T, C_T\}$ are parameters that is tangent to the vehicle trajectory. With this two equations, the normal vector of the plane $\{A', B', C'\}$ can be determined uniquely. Equation (8) defines the position of the plane, that is in accordance with the nearest occupied grid in y dimension. $d(P', V)$ measures the distance between the plane we want to get P' and the vehicle V . With this equation, parameter D' can be gotten. Thus the parameters of the projection plane $\{A', B', C', D'\}$ are fully determined. The value in every grid of the projection plane depends on the values in the 3D grids that are projected on it (*i.e.* $\mathbf{P}(n|\mathbf{Q})$). Here we take a weight for the distance the grid to the road center. The value in the 2D grid is:

$$\mathbf{P}(n) = \sum_{i=1}^k W(f(c)_i) \mathbf{P}_i \quad (9)$$

In equation (9), $W(f(c)_i)$ is the weight of i th grid and $W(f(c)_i)$ is $\propto \frac{1}{f(c)}$. \mathbf{P}_i is the corresponding possibility of the i th grid. With these procedures, the compressed road scene map is generated and it is vividly shown in Fig.4.

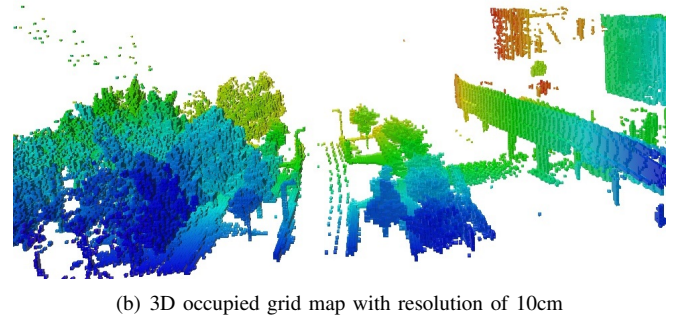
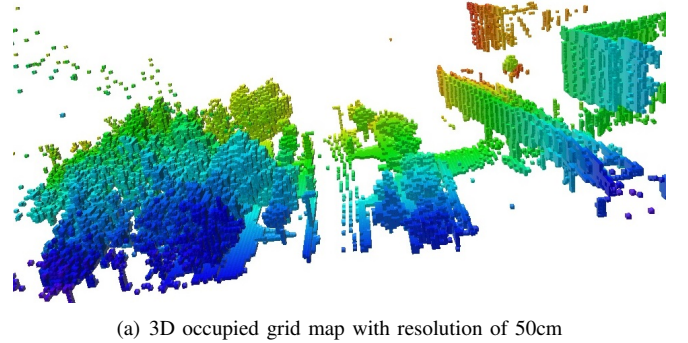
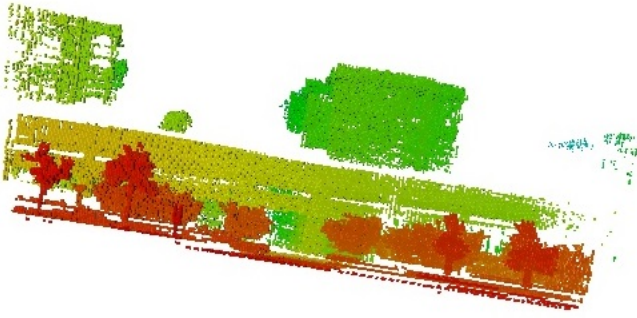


Fig. 3. The 3D occupied grid map with different resolutions. (a) The 50cm resolution 3D occupied grid map of the point cloud shown in Fig.2. (b) The 10cm resolution 3D occupied grid map of the point cloud shown in Fig.2.

IV. LOCALIZATION

In this section, the localization algorithm will be presented in detail. The localization is a matching process between the compressed road scene map and the live data. Usually,



(a) Road DNA of the scene in Fig.2.



(b) A binary occupied grid map of (a)

Fig. 4. Road DNA of the same scene in Fig.2 and Fig.3. (a) The colored road scene map (different colors denote different distance from the center of the road). (b) A binary presentation of (a).

a cost function that computes the similarity of two data sets is adopted, then optimization is utilized to best align the two data sets. Finally a transformation matrix can be gotten. Because of the property of compressed road scene map (similar to gray-level based image), the Normalized Information Distance (NID) serves as the matching method in this paper, as it can measure the similarity between two image-like source well. NID is an entropy based method, which is often used for registering data from multi-modal sources. In [11] and [12], it is successfully used to evaluate similarity of two images. Given two measurements A and B , NID is defined as:

$$NID(A, B) = \frac{H(A, B) - MI(A, B)}{H(A, B)} \quad (10)$$

Here, H and MI represent the Shannon entropy and mutual information over the variable:

$$H(A) = -\sum_{a \in A} p(a) \log p(a) \quad (11)$$

$$H(A, B) = -\sum_{a \in A} \sum_{b \in B} p(a, b) \log p(a, b) \quad (12)$$

$$MI(A, B) = H(A) + H(B) - H(A, B) \quad (13)$$

Given a transformation relation $G = (x, y, \theta)$, the matching process is to find a \hat{G} that minimize the NID between compressed road scene map and live data:

$$\hat{G} = \arg \min_G NID(\mathbf{Q}(G), \mathbf{M}) \quad (14)$$



Fig. 5. The platform used for experiments in this paper. In the red circle is the LIDAR used in this system, *i.e.*, Velodyne VLP-16.

In equation (14), $\mathbf{Q}(G)$ is the live data transformed by G , \mathbf{M} is the map loaded. In this paper, we also integrate information from IMU, odometer and GPS like [2]. Given the live data \mathbf{Q}_k , the loaded map \mathbf{M} , the localization can be treated as a maximum posteriori estimation problem as follows:

$$\hat{\mathbf{x}}_k = \arg \max_{\mathbf{x}_k} p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{z}_k, \mathbf{Q}_k, \mathbf{M}) \quad (15)$$

In equation (15), \mathbf{x}_k and \mathbf{x}_{k-1} are the pose and position of the vehicle at time interval k and $k-1$, respectively. \mathbf{u}_k is the vehicle movement in time interval k measured by IMU and odometer, \mathbf{z}_k is GPS if the signal is available. \mathbf{Q}_k and \mathbf{M} are the live road scene flattening and compressed road scene, respectively. Applying Bayes' rule:

$$p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{z}_k, \mathbf{Q}_k, \mathbf{M}) \propto p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{u}_k, \mathbf{z}_k) p(\mathbf{Q}_k | \mathbf{x}_k, \mathbf{M}) \quad (16)$$

In equation (16), the first term is computed by dead reckoning, the second term is computed by matching method:

$$p(\mathbf{Q}_k | \mathbf{x}_k, \mathbf{M}) \propto 1/NID(\mathbf{Q}_k(\mathbf{x}_k), \mathbf{M}) \quad (17)$$

For real-time application, we exploit a graphics processing unit (GPU) here, which will be presented in detail together with the experiments and results analysis in the next section.

V. EXPERIMENTAL RESULTS

In this section we will present our experimental approach to verify long term high-precision localization without big errors. The experiment is done in the campus of Shanghai Jiao Tong University which contains most of urban scenes (*i.e.*, structured and unstructured). The experimental platform is SJTU CyberTiggo shown in Fig.5. A multi-beam laser

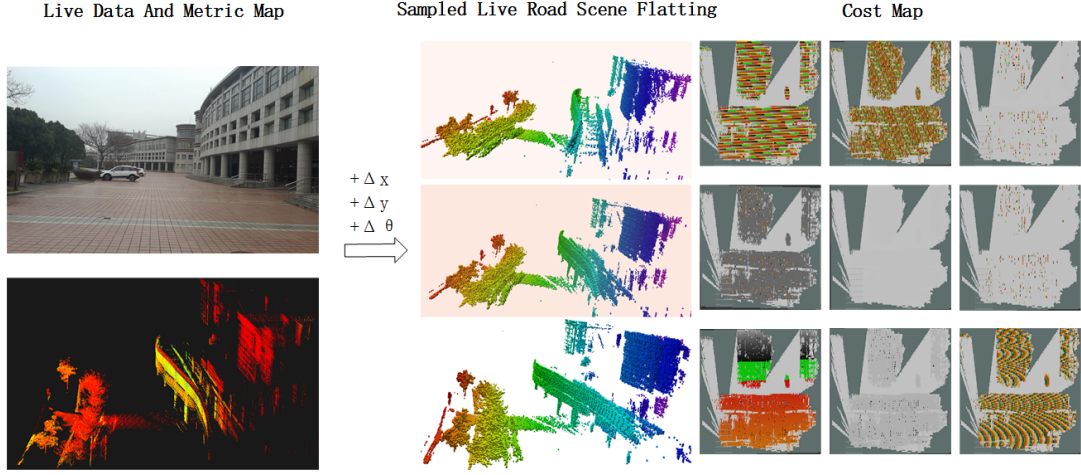


Fig. 6. NID based registration process. First, live data are collected by laser scanner. The road scene flatting is sampled at a number of offsets $x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n, \theta_1, \theta_2, \dots, \theta_n$. Then the samples are matched against the compressed road scene map which yields a number of NMI values (presented as cost maps). The minimum is selected as the matching result.

TABLE I
LOCALIZATION ALGORITHM PARAMETERS

Symbol	Parameter	Value
$l(\mathbf{M})$	map piece length	50m
l	Resolution	0.5m, 0.25m, 0.05m
s	Size of map	30MB
$s(p)$	Size of raw point cloud	50GB
f	Output frequency	10Hz
n	Number of x, y, θ offset	10
d	Effective range of laser scan	30m

scanner is mounted in front of the vehicle vertically. Low-cost GPS, Xsens-MTi, two encoders are equipped to provide data of movement observation.

The experimental dataset consists of five traversals of an approximately 20km route. The traversals were made at different times of day. In some areas, the GPS signals were in bad condition, while our localization performed well for autonomous driving. Besides that, the GPS we used is common GPS with error of 5-10 meters, it mainly did benefit of initializing. The localization error by our method is less than 1 meter in most areas, which indicates the improvement is due to Road DNA based localization, while not due to GPS. The ground truth was estimated by RTK-GPS. Table I lists parameters used for the localization algorithm. In the table, it can be seen that the size of Shanghai Jiao Tong University map is 30MB. Comparing to the size of point cloud map in the same place-50GB, the compressed road scene is much smaller, which indicates that this type of map is suitable for large-scale mapping and localization.

A. NID based registration

First, NID based registration between live road scene flatting and metric map is evaluated. The road scene flatting is sampled at a number of offsets in lateral, longitude and heading orientation. Then every sample aligns with the metric map loaded from map manager by NID. After that,

TABLE II
LOCALIZATION PERFORMANCE DURING EXPERIMENT

Parameters	MAE	Max	90% CI
Lateral(x)	0.35m	0.49m	0.30m
Longitudinal(y)	0.40m	1.0m	0.40m
Heading(θ)	0.73°	2.57°	1.00°

the best matching one can be confirmed by selecting the matching pair which has the lowest NID value. This sampling method requires a significant computational cost. Since the samples are independent, it is possible to exploit modern GPU processors to simultaneously evaluate a large number of observation likelihoods in parallel. So we exploit GPU pipeline to accelerate the sampling matching process. A series of the matching examples are illustrated in Fig.6, where different colors denote different NID value, the darker the color, the bigger the NID value.

The compressed road scene map has a capacity of dynamic objects tolerance. In Fig.6, there are a moving car and some pedestrian in the live data. After NID matching, the transformation relation can be still estimated correctly as is shown in the figure.

B. Localization results

Fig.7 shows the localization error in lateral, longitudinal and heading orientation. It can be seen that the biggest error in the lateral and longitudinal direction is less than 1m, the biggest heading error is less than 3°. With precision in this level, the autonomous vehicle can travel freely and safely in most cases. In the figure, the localization errors by filtered dead reckoning and by point cloud matching method are also presented. Localization results by trajectory filter is initialized by high-precision GPS, clearly it suffers from accumulative errors as the travel distance increases. In this paper, localization by Iterative Closest Point (ICP) algorithm serves as comparison. In comparison experiments,

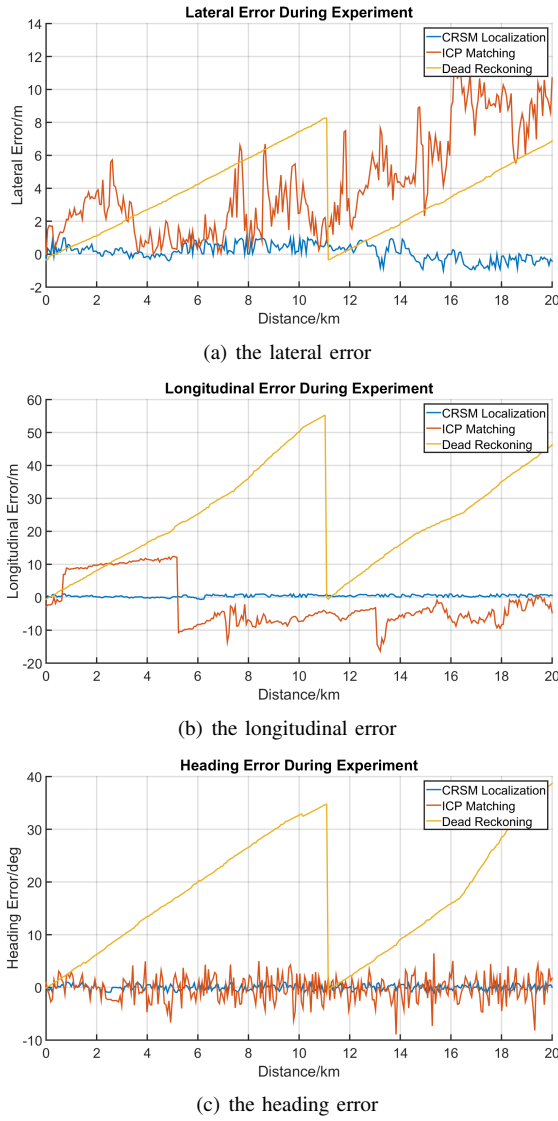


Fig. 7. Errors during experiment, the ground truth is produced by RTK-GPS. (a) The lateral error. (b) The longitudinal error. (c) The heading error. Three localization methods are compared here, *i.e.*, CRSM (Compressed Road Scene Map) based localization (method in this paper), ICP matching based localization and dead reckoning. It is a 20 kilometers long experiment, the errors in lateral and longitudinal directions are both in 1m. And the heading error is in 1 degree. While the ICP based method is not robust and not real-time and dead reckoning method suffers from accumulative error badly. After traveled for 11km, the dead reckoning was re-initialized as shown.

ICP directly matches the live point cloud with the prior-built point cloud map, while the point cloud was not converted to Road DNA and flattening road scene. Localization by ICP can hardly be done in real time. Besides that, it may be trapped in local minimum or be easily affected by kidnapping problem. After traveled for 11km, the vehicle stopped and the dead reckoning was re-initialized. So in the figure, the errors by dead reckoning become around zero again. In this traversal, the vehicle traveled most parts of the campus of Shanghai Jiao Tong University. Table II summarizes the localization performance of experiments through five months.

The mean absolute error (MAE) is 0.35m, 0.40m and 0.73° in lateral, longitudinal and heading, respectively. And most of the localization errors (90%) fell in 0.30m, 0.40m and 1.0° .

VI. CONCLUSION

In this paper, a new type of map (*i.e.*, Road DNA) is presented to solve the high-precision localization problem for intelligent vehicles in large-scale area robustly. The new type of map (Road DNA) is much more storage saving comparing to the traditional map (*i.e.*, point cloud map, vision map), which gives it a promising application in large-scale (like city-level) use. Besides that, the combination of NMI and GPU makes the map matching process precise and real-time. The proposed method shows better performance than widely used ICP method. Experiments show the precision, robustness and validation of this localization framework. In the future work, we will test and verify its validation in city-level (*e.g.* Shanghai) use. And the scheme of Road DNA update and 3D cartography (methods that mitigate other sensor data), approach of tightly-coupled GNSS and 3D sensor will be also exploited.

REFERENCES

- [1] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," in *Autonomous robot vehicles*, pp. 167–193, Springer, 1990.
- [2] W. Maddern, G. Pascoe, and P. Newman, "Leveraging Experience for Large-Scale LIDAR Localisation in Changing Cities," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, (Seattle, WA, USA), May 2015.
- [3] R. W. Wolcott and R. M. Eustice, "Fast LIDAR localization using multiresolution Gaussian mixture maps," in *Proceedings of the IEEE International Conference on Robotics and Automation*, (Seattle, WA, USA), pp. 2814–2821, May 2015.
- [4] Z. Chong, B. Qin, T. Bandyopadhyay, M. H. Ang, E. Frazzoli, and D. Rus, "Synthetic 2d lidar for precise vehicle localization in 3d urban environment," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pp. 1554–1559, IEEE, 2013.
- [5] H. Latgahn and C. Stiller, "Vision-only localization," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 15, no. 3, pp. 1246–1257, 2014.
- [6] C. Linegar, W. Churchill, and P. Newman, "Work Smart, Not Hard: Recalling Relevant Experiences for Vast-Scale but Time-Constrained Localisation," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, (Seattle, WA, USA), May 2015.
- [7] Z. Alsayed, G. Bresson, F. Nashashibi, and A. Verroust-Blondet, "Pmlslam: a solution for localization in large-scale urban environments," in *PPNIV-IROS 2015*, 2015.
- [8] J. Zumberge, M. Heflin, D. Jefferson, M. Watkins, and F. H. Webb, "Precise point positioning for the efficient and robust analysis of gps data from large networks," *Journal of Geophysical Research: Solid Earth (1978–2012)*, vol. 102, no. B3, pp. 5005–5017, 1997.
- [9] R. Schnabel, R. Wahl, and R. Klein, "Efficient ransac for point-cloud shape detection," in *Computer graphics forum*, vol. 26, pp. 214–226, Wiley Online Library, 2007.
- [10] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: An efficient probabilistic 3D mapping framework based on octrees," *Autonomous Robots*, 2013.
- [11] 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.
- [12] R. W. Wolcott and R. M. Eustice, "Visual localization within LIDAR maps for automated urban driving," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, (Chicago, IL, USA), pp. 176–183, September 2014.