

Long-Term Urban Vehicle Localization Using Pole Landmarks Extracted from 3-D Lidar Scans

Alexander Schaefer, Daniel Büscher, Johan Vertens, Lukas Luft, Wolfram Burgard

Abstract—Due to their ubiquity and long-term stability, pole-like objects are well suited to serve as landmarks for vehicle localization in urban environments. In this work, we present a complete mapping and long-term localization system based on pole landmarks extracted from 3-D lidar data. Our approach features a novel pole detector, a mapping module, and an online localization module, each of which are described in detail, and for which we provide an open-source implementation [1]. In extensive experiments, we demonstrate that our method improves on the state of the art with respect to long-term reliability and accuracy: First, we prove reliability by tasking the system with localizing a mobile robot over the course of 15 months in an urban area based on an initial map, confronting it with constantly varying routes, differing weather conditions, seasonal changes, and construction sites. Second, we show that the proposed approach clearly outperforms a recently published method in terms of accuracy.

I. INTRODUCTION

Intelligent vehicles require accurate and reliable self-localization systems. Accurate, because an exact pose estimate enables complex functionalities such as automatic lane following or collision avoidance in the first place. Reliable, because the quality of the pose estimate must be maintained independently of environmental factors in order to ensure safety.

Satellite-based localization systems like RTK-GPS or DGPS seem to be an efficient solution, since they achieve centimeter-level accuracy out of the box. However, they lack reliability. Especially in urban areas, buildings that obstruct the line of sight between the vehicle and the satellites can decrease accuracy to several meters [2], [3]. Localization on the basis of dense maps like grid maps, point clouds, or polygon meshes represents a more reliable alternative [4]. On the downside, dense approaches require massive amounts of memory that quickly become prohibitive for maps on larger scales. This is where landmark maps come into play: By condensing billions of raw sensor data points into a comparably small number of salient features, they can decrease the memory footprint by several orders of magnitude [5].

In this work, we present an approach to long-term 2-D vehicle localization in urban environments that relies on pole landmarks extracted from mobile lidar data. Poles occur as parts of street lamps, traffic signs, as bollards and tree

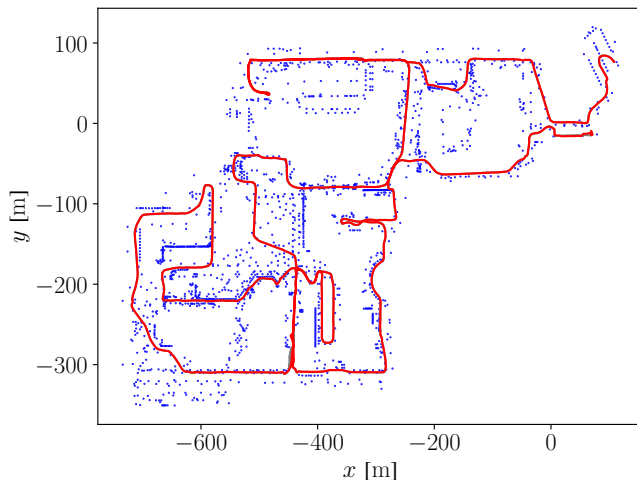


Fig. 1: Pole landmark map created from the NCLT dataset [3] and trajectory of an experimental run 15 months after map creation. The blue dots represent the landmarks. The gray line corresponds to the ground-truth trajectory. Most of it is covered by the red line, which represents the estimate produced by the presented method. The mean position difference between both trajectories, formally defined in section IV-A, amounts to 0.31 m.

trunks. They are ubiquitous in urban areas, long-term stable and invariant under seasonal and weather changes. Since their geometric shape is well-defined, too, poles are well suited to serve as landmarks that enable accurate and reliable localization.

Our localization process is subdivided into a mapping and a localization phase. During mapping, we use the pole detector presented below to extract pole landmarks from lidar scans and register them with a global map via a given ground-truth vehicle trajectory. During localization, we employ a particle filter to estimate the vehicle pose by aligning the pole detections from live sensor data with those in the map. Figure 1 shows an exemplary localization result.

We are not the first ones to propose this kind of localization technique: The next section provides an overview over the numerous related works. However, to the best of our knowledge, we are the first ones to present a pole detector that does not only consider the laser ray endpoints, but also the free space in between the laser sensor and the endpoints, and to demonstrate reliable and accurate vehicle localization based on a map of pole landmarks on large time scales. While related works usually evaluate localization performance on

This work has been partially supported by Samsung Electronics Co. Ltd. under the GRO program.

All authors are with the Department of Computer Science, University of Freiburg, Germany.

{aschaefer, buescher, vertensj, luft, burgard}
@cs.uni-freiburg.de

978-1-7281-3605-9/19/\$31.00 ©2019 IEEE

a short sample trajectory of at most a few minutes length, we successfully put our approach to the test on a publicly available long-term dataset that contains 35 hours of data recorded over the course of 15 months – including varying routes, construction zones, seasonal and weather changes, and lots of dynamic objects. Additional control experiments show that the presented method is not only reliable, but significantly outperforms a recently published state-of-the-art approach in terms of accuracy, too.

II. RELATED WORK

In recent years, a number of authors have addressed the specific question of vehicle localization via pole landmarks extracted from lidar scans. Any solution to this question consists of at least two parts: a pole detector and a landmark-based pose estimator. The detector developed by Weng et al. [6], for example, tessellates the space around the lidar sensor and counts the number of laser reflections per voxel. Poles are then assumed to be located inside contiguous vertical stacks of voxels that all exceed a reflection count threshold. In order to extract the pole parameters from these clusters, the detector fits a cylinder to all points in a stack via RANSAC [7]. For 2-D pose estimation, the authors employ on a particle filter with nearest-neighbor data association. Sefati et al. [8] present a pole detector that removes the ground plane from a given point cloud, projects the remaining points onto a horizontal regular grid, clusters neighboring cells based on occupancy and height, and fits a cylinder to each cluster. Like Weng et al., Sefati et al. obtain their 2-D localization estimate from a particle filter that performs nearest-neighbor data association. Kümmerle et al. [5] make use of Sefati et al.’s pole detector, but to further refine the localization estimate, they also fit planes to building façades in the laser scans and lines to lane markings in stereo camera images. Like the above works, their pose estimator relies on a Monte Carlo method to solve the data association problem, but uses optimization to compute the most likely pose. More specifically, in the data association stage, it builds a local map by accumulating the landmarks detected over the past timesteps based on odometry. It then samples multiple poses around the current GPS position, uses these pose hypotheses to project the local map into the global map, and identifies the most probable hypothesis via a handcrafted landmark matching metric. Given the resulting data associations, it refines the current vehicle pose estimate via nonlinear least squares optimization over a graph of past vehicle poses and landmarks.

Spangenberg et al. [9] extract pole landmarks not from lidar scans, but from stereo camera images. In order to estimate the vehicle pose, they feed wheel odometry, GPS data, and online pole detections to a particle filter.

While the approaches above all provide a complete localization system consisting of a pole extractor and a landmark-based localization module, there exist a variety of research papers that focus solely on pole extraction. Extracting poles from lidar data is a common problem in road infrastructure maintenance and urban planning. In this domain, researchers

are not only interested in fitting geometric primitives to the data and determining pole coordinates, but also in precise point-wise segmentation. Brenner [10], Cabo et al. [11], Tombari et al. [12], and Rodriguez et al. [13] present different methods to extract pole-like objects from point clouds, i.e. without accounting for free space information. The approaches of Yu et al. [14] and Wu et al. [15] specifically target street lamp poles, while Zheng et al. [16] provide a solution to detect poles that are partially covered by vegetation. Yokoyama et al. [17] not only extract poles, but they classify them as lamp posts, utility poles, and street signs. Ordóñez et al. [18] build upon the pole detector proposed by Cabo et al. [11] and classify the results into six categories, including trees, lamp posts, traffic signs, and traffic lights. Li et al. [19] take classification one step further by decomposing multifunctional structures, for example a light post carrying traffic signs, into individual elements.

Poles are not the only landmarks suitable for vehicle localization. Qin et al. [20] investigate Monte Carlo vehicle localization in urban environments based on curb and intersection features. As demonstrated by the works of Schindler [21] and Schreiber et al. [22], road markings as landmarks can also yield high localization accuracy. Hata and Wolf [23] feed both curb features and road markings to their particle filter. Welzel et al. [24] explore the idea of using traffic signs as landmarks. Although traffic signs occur less frequently in urban scenarios compared to other types of road furniture like road markings or street lamp poles, they offer the advantage of not only encoding a position, but also an unambiguous ID. Finally, Im et al. [25] explore urban localization based on vertical corner features, which appear at the corners of buildings, in monocular camera images and lidar scans.

III. APPROACH

The proposed 2-D vehicle localization system consists of three modules: the pole extractor, the mapping module, and the localization module. During the initial mapping phase, the pole extractor reduces a given set of lidar scans to pole landmarks. The mapping module then uses the ground-truth sensor poses to build a global reference map of these landmarks. During the subsequent localization phase, the pole extractor processes live lidar data and passes the resulting landmarks to the localization module, which in turn generates a pose estimate relative to the global map. In the following, we detail each of these modules and their interactions.

A. Pole Extraction

The pole extraction module takes a set of registered 3-D lidar scans as input and outputs the 2-D coordinates of the centers of the detected poles with respect to the ground plane, along with the estimated pole widths. To that end, it builds a 3-D occupancy map of the scanned space, applies a pole feature detector to every voxel, and regresses the resulting pole map to a set of pole position and width estimates.

To describe these three steps mathematically, we denote a single laser measurement – a ray – by $z := \{u, v\}$, where u and v represent its Cartesian starting point and endpoint,

respectively. All measurements $Z := \{z_i\}$ are assumed to be registered with respect to the map coordinate frame, whose x - y plane is aligned with the ground plane. The measurements can be taken at different points in time, but the timespan between the first and the last measurement needs to be sufficiently small in order not to violate our assumption that the world is static. Now, we tessellate the map space, trace the laser rays, and model the posterior probability that the j -th voxel reflects an incident laser ray according to Luft et al. [26] by

$$p(\mu_j | Z) = \text{Beta}(h_j + \alpha, m_j + \beta).$$

Here, h_j and m_j denote the numbers of laser reflections and transmissions in the j -th cell, whereas α and β are the parameters of the prior reflection probability $p(\mu_j) = \text{Beta}(\alpha, \beta)$, which we determine in accord with [26] by

$$\alpha = -\frac{\gamma(\gamma^2 - \gamma + \delta)}{\delta}, \quad \beta = \frac{\gamma - \delta + \gamma\delta - 2\gamma^2 + \gamma^3}{\delta},$$

where $M := \{h_j(h_j + m_j)^{-1}\}$ denotes the maximum-likelihood reflection map, and where $\gamma := \mathbb{E}[M]$, $\delta := \text{var}[M]$ represent its mean and variance, respectively. Please note that $\{p(\mu_j | Z)\}$ is a full posterior map: In contrast to M , which assigns each voxel the most probable reflection rate, it yields a posterior distribution over every reflection rate possible.

Since we want to extract poles based on occupancy probability, not on reflection rate, we convert $\{p(\mu_j | Z)\}$ to an occupancy map $O := \{o_j\}$. Assuming that a cell is occupied if its reflection rate exceeds a threshold μ_o , we formulate the occupancy probability by integration:

$$o_j := \int_{\mu_o}^1 p(\mu_j | Z) d\mu_j.$$

Next, a pole feature detector transforms O to a 2-D map of pole scores S in the ground plane. Each pixel of S encodes the probability that a pole is present at the corresponding location. The transformation from O to S follows a set of heuristics that are based on the definition of a pole as a vertical stack of occupied voxels with quadratic footprint, laterally surrounded by a hull of free voxels. First, we create a set of intermediate 3-D score maps of the same size as O , each denoted by $Q_a := \{q_{a,j}\}$. Every cell $q_{a,j}$ tells how probable it is that this portion of space is part of a pole with edge length a , where $a \in \mathbb{N}^+$ is measured in units of grid spacing:

$$q_{a,j} := \max_{k \in \text{inside}(j,a)} \left(\frac{\sum_{l \in \text{inside}(k,a)} o_l}{a^2} - \max_{l \in \text{outside}(k,a,f)} o_l \right).$$

Here, $\text{inside}(j, a)$ and $\text{outside}(j, a, f)$ are functions that, given a map index j and a pole width a , return a set of indices into voxels in the same horizontal map slice as j . While the former outputs the indices of all voxels inside the pole, the latter returns the indices corresponding to the supposedly free region around the pole with thickness $f \in \mathbb{N}^+$. Both functions assume that the lower left lateral walls of the pole are aligned with the lower left lateral sides of the j -th

voxel. With these definitions, the argument of the enclosing maximum operator amounts to the difference between the mean occupancy value inside the pole and the maximum occupancy value of the volume of free space around the pole. The resulting score lies in the interval $[-1, 1]$: the higher the score, the greater the probability that the corresponding partition of space is part of a pole. Second, we regress from the resulting 3-D maps $\{Q_a\}$ to 2-D by merging them into a single map $Q := \{q_j\} := \{\max_a q_{a,j}\}$ and by determining for each horizontal position in Q the contiguous vertical stack of voxels that all surpass a given score threshold q_{\min} . After discarding all stacks that fall below a certain height threshold h_{\min} and computing the mean score for each of the remaining stacks, we obtain the desired 2-D score map S .

Finally, we convert this discrete score map to a set of continuous pole position and width estimates. We identify the pole positions as the modes of S , which we determine via mean shift [27] with a Gaussian kernel and with the local maxima of S as seed points. The width estimate of each pole is computed as the weighted average over all pole widths a , where for every a , the weight is the mean of all cells in Q_a that touch the pole.

The presented algorithm differs from other pole extractors in the fact that it is based on ray tracing. By considering not only the scan endpoints, but also the starting points, it explicitly models occupied and free space. In contrast, most other methods assume the space around the sensor to be free as long as it does not register any reflections. The absence of reflections, however, can have two reasons: The respective region is in fact free, or the lidar sensor did not cover region due to objects blocking its line of sight or its limited range.

B. Mapping

In theory, the global reference map could be built by simply applying the pole extractor to a set of registered laser scans that cover the area of interest. In practice, the high memory complexity of grid maps and laser scans often renders this naive approach infeasible. To create an arbitrarily large landmark map with limited memory resources, we partition the mapping trajectory into shorter segments of equal length and feed the lidar measurements taken along each segment to the pole extractor one by one. For the sake of consistency, we take care that the intermediate local grid maps are aligned with the axes of the global map and that all of them have the same raster spacing. The intermediate maps, whose sizes are constant and depend on the sensor range, usually fit into memory easily. Processing all segments provides us with a set of pole landmarks. If the length of a trajectory segment is smaller than the size of a local map, the local maps overlap, a fact that can lead to multiple landmarks representing a single pole. In order to merge these ambiguous landmarks, we project all poles onto the ground plane, yielding a set of axis-aligned squares. If multiple squares overlap, we reduce them to a single pole estimate by computing a weighted average over their center coordinates and widths. Each weight equals the mean pole score, which we determine by averaging over the scores of all voxels that

touch the pole in all score maps Q_a . If there is no overlap, we integrate the corresponding pole into the global reference map without further ado.

As a side benefit, this mapping method allows us to filter out dynamic objects at the landmark level using a sliding-window approach: A local landmark is integrated into the reference map only if it was seen at least c times in the past w local maps, where $c \leq w$; $c, w \in \mathbb{N}^+$. Correspondences between landmarks are again determined via checking for overlapping projections in the ground plane.

C. Localization

During online localization, we continuously update the vehicle pose based on the collected odometry measurements and periodically correct the estimate by matching online pole landmarks, which we extract from the most recent local map, against the reference map. We build the local map by accumulating laser scans along a segment of the trajectory and by registering them via odometry. To filter out dynamic objects, we apply the sliding-window approach described in the previous section.

A particle filter is well suited for the localization task [28], because it can not only maintain multiple pose hypotheses in parallel, but also handle global localization. At time t , each particle corresponds to a 2-D vehicle pose hypothesis, represented by the 3×3 homogeneous transformation matrix X_t . To perform the motion update, we assume Gaussian motion noise Σ and sample from a trivariate normal distribution in χ :

$$X_t = \text{transform}(\xi) X_{t-1} \quad | \quad \xi \sim \mathcal{N}(\chi, \Sigma),$$

where $\chi := [x, y, \phi]^\top$ denotes the latest relative odometry measurement, with x, y , and ϕ representing the translation and the heading of the vehicle, respectively. The function $\text{transform}([x, y, \phi]^\top)$ converts the input vector to the corresponding 3×3 transformation matrix. In each measurement update, we determine the data associations between the online landmarks $\Lambda := \{\lambda_k\}$ and the landmarks in the reference map $L := \{l_n\}$ via nearest-neighbor search in a k -D tree, assume independence between the elements of Λ , and update the particle weights according to the measurement probability

$$p(\Lambda | X, L) = \prod_k p(\lambda_k | X, l_{n(k)}),$$

where $n(k)$ is the data association function that tells the index of the reference landmark associated with the k -th online landmark. To evaluate the above equation, we need to define a measurement model

$$p(\lambda_k | X, l_{n(k)}) := \mathcal{N}(\|X\lambda_k - l_{n(k)}\|, \sigma) + \epsilon,$$

with the reference and online landmarks represented by homogeneous 2-D position vectors $[x, y, 1]^\top$, and where we assume isotropic position uncertainty σ of the reference landmarks. The constant addend $\epsilon \in \mathbb{R}^+$ accounts for the probability of discovering a pole that is not part of the map. This probability can be estimated by generating a

global map from one run, generating a set of local maps from data recorded on the same trajectory in a second run, and computing the numbers of matched and unmatched landmarks.

IV. EXPERIMENTS

In order to evaluate the proposed localization system, we perform two series of experiments. The complete implementation is publicly available [1]. In the first series, we assess the system's long-term localization reliability and accuracy on the NCLT dataset [3]. While these experiments provide profound insights into the performance of the developed method, the results are absolute and do not allow direct comparisons with other methods, because to the best of our knowledge, we are the first to test landmark-based localization on NCLT. For this reason, we base the second experiment series on the KITTI dataset [29]. That allows us to repeat the experiments performed by the authors of another state-of-the-art localization method, only that this time, we use the system presented above.

A. Localization on the NCLT Dataset

The NCLT (North Campus Long-Term) dataset [3] was acquired with a two-wheeled Segway robot on one of the campuses of the University of Michigan, USA. The data is perfectly suited for testing the capabilities of any system that targets long-term localization in urban environments: Equipped with a Velodyne HDL-32E lidar, GPS, IMU, wheel encoders, and a gyroscope, among others, the robot recorded 27 trajectories with an average length of 5.5 km and an average duration of 1.3 h over the course of 15 months. The recordings include different times of day, different weather conditions, seasonal changes, indoor and outdoor environments, lots of dynamic objects like people and moving furniture, and two large construction projects that evolve constantly. Although the routes differ significantly between sessions, the trajectories have a large overlap.

The main difference between NCLT and the data used to evaluate all other pole-based localization methods we surveyed lies in its extent: While related works briefly demonstrate the plausibility of their approaches by evaluating localization performance on datasets with durations between 46 s and 30 min, we focus on long-term reliability and accuracy and process 35 h of data spread over more than one year.

Before localizing, we build a reference map of the poles on the campus. To that end, we feed the laser scans and the ground-truth robot poses of the very first session to our mapping module. Unfortunately, the ground truth provided by NCLT is not perfect. It consists of optimized poses spaced in intervals of 8 m, interpolated by odometry. Consequently, point clouds accumulated over a few meters exhibit considerable noise, as illustrated in figure 2. For that reason, we set the distance of the trajectory segments to build local maps to 1.5 m, the raster spacing of the grid maps to 0.2 m, and the occupancy threshold to $\mu_o = 0.2$. During mapping and localization, the pole extractor discards all poles below a

Session Date	f_{map} [%]	Δ_{pos} [m]	RMSE $_{\text{pos}}$ [m]	Δ_{ang} [$^{\circ}$]	RMSE $_{\text{ang}}$ [$^{\circ}$]
2012-01-08	100.0	0.130	0.178	0.663	0.857
2012-01-15	8.5	0.156	0.225	0.760	0.999
2012-01-22	5.1	0.172	0.222	0.939	1.291
2012-02-02	0.4	0.155	0.205	0.720	0.975
2012-02-04	0.1	0.144	0.195	0.684	0.903
2012-02-05	0.5	0.148	0.210	0.691	0.947
2012-02-12	0.8	0.269	1.005	0.802	1.040
2012-02-18	0.8	0.149	0.221	0.699	0.938
2012-02-19	0.0	0.148	0.194	0.704	0.944
2012-03-17	0.0	0.149	0.191	0.830	1.062
2012-03-25	0.0	0.200	0.262	1.418	1.836
2012-03-31	0.0	0.143	0.184	0.746	0.973
2012-04-29	0.0	0.170	0.251	0.829	1.079
2012-05-11	5.5	0.161	0.225	0.773	0.998
2012-05-26	0.4	0.158	0.217	0.690	0.889
2012-06-15	0.4	0.180	0.238	0.659	0.879
2012-08-04	0.3	0.210	0.340	0.884	1.143
2012-08-20	3.8	0.189	0.264	0.711	0.941
2012-09-28	0.3	0.206	0.311	0.731	0.952
2012-10-28	1.4	0.217	0.338	0.693	0.919
2012-11-04	2.5	0.257	0.456	0.746	0.996
2012-11-16	2.7	0.403	0.722	1.467	2.031
2012-11-17	0.4	0.243	0.377	0.686	0.959
2012-12-01	0.0	0.266	0.492	0.674	0.930
2013-01-10	0.0	0.217	0.278	0.689	0.911
2013-02-23	0.0	2.470	5.480	1.083	1.769
2013-04-05	0.0	0.365	0.920	0.654	1.028

TABLE I: Results of our experiments with the NCLT dataset, averaged over ten localization runs per session. The variables Δ_{pos} and Δ_{ang} denote the mean absolute errors in position and heading, respectively, RMSE $_{\text{pos}}$ and RMSE $_{\text{ang}}$ represent the corresponding root mean squared errors, while f_{map} denotes the fraction of lidar scans per session used to build the reference map.

minimum pole height of $h_{\text{min}} = 1$ m and below a minimum pole score of $q_{\text{min}} = 0.6$. The extent of the local maps is chosen $30 \text{ m} \times 30 \text{ m} \times 5 \text{ m}$ in x , y , and z of the map frame, respectively. Figure 3 illustrates the corresponding results.

Although the first session covers most of the campus, the robot occasionally roams into unseen regions during later sessions. For that reason, we iterate over all subsequent sessions, too, but add landmarks to the global map only if the corresponding laser scans are recorded at a minimum distance of 10 m from all previously visited poses. Table I shows that after the second session, the fractions of scans per session that contribute to the map drop to $f_{\text{map}} \leq 5.5\%$.

During localization, odometry mean and covariance estimates are generated by fusing wheel encoder readings, gyroscope, and IMU data in an extended Kalman filter. The particle filter contains 5000 particles, which we initialize by uniformly sampling positions in a circle with radius 2.5 m around the earliest ground-truth pose. The headings are uniformly sampled in $[-5^{\circ}, 5^{\circ}]$. To maximize reliability, we inflate the motion noise by a factor of four, which corresponds to doubled standard deviation, define the position uncertainty of the poles in the global map as $\sigma = 1 \text{ m}^2$,

and set the addend in the measurement probability density to $\epsilon = 0.1$. We resample particles whenever the number of effective particles $n_{\text{eff}} := (\sum_i w_i^2)^{-1} < 0.5$, where w_i is the weight of the i -th particle, via low-variance resampling as described by Thrun et al. [28]. In order to obtain the pose estimate, we select the best 10% of the particles and compute the weighted average of their poses.

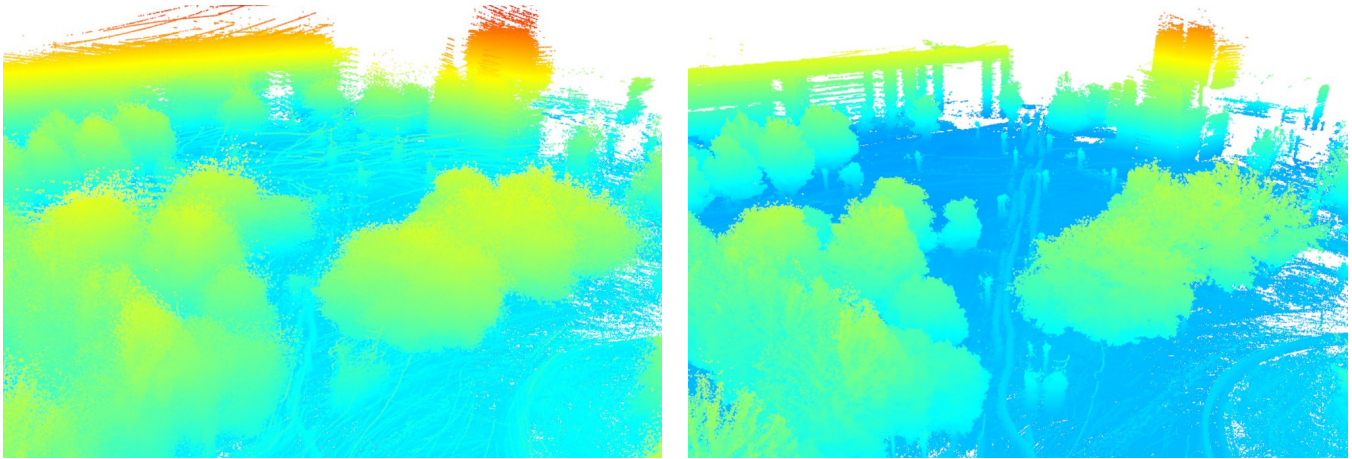
Table I presents for each of the 27 sessions the corresponding position and heading errors. To generate these values, we run the localization module ten times per session, evaluate the deviation of our estimate from ground truth every 1 m along the ground-truth trajectory, compute the means and RMSEs, and average these metrics over the ten sessions. The results demonstrate that the proposed method achieves both high reliability and accuracy, even if the data used for mapping and for localization lie 15 months apart: The particle filter never even partially diverges, except for one late session discussed below. Furthermore, despite the inaccuracies in ground truth, which affect both the global map and the evaluation, it achieves a mean positioning accuracy over all sessions of 0.284 m. Looking at the evolution of the errors over time, we observe slightly increasing magnitudes. This is due to changes in campus infrastructure accumulating over time and rendering the initial map more and more outdated.

In session 2012-02-23, these changes eventually cause the localization module to temporarily lose track of the exact robot position. The diverging behavior reproducibly occurs when the robot drives along a row of construction barrels that fence a large construction site. When the global map was built, these barrels were located on the footpath. Just before the session in question, however, the barrels were moved laterally by a few meters, while maintaining their longitudinal positions. Since the barrels are the only landmarks in the corresponding region, the localizer “corrects” the robot position so that the incoming pole measurements match the map. Having passed the construction site, the localizer is confident about its wrong position estimate, which is why it takes some time until the particle cloud diverges and the robot relocates. The positioning error over all sessions except 2012-02-23 amounts to 0.200 m.

Lastly, we describe the runtime requirements of our method stochastically. On a 2011 quad-core PC with dedicated GPU, we measure an average 1.33 s for pole extraction with our open-source Python implementation [1], which corresponds to processing 0.5 million laser data points per seconds. The measurement step with data association requires a mean computation time of 0.09 s. These two steps pose by far the highest computational requirements and make others, like the measurement update, negligible.

B. Localization on the KITTI Dataset

As delineated in section II, Kümmerle et al. [5], Weng et al. [6], and Sefati et al. [8] present methods for vehicle localization with pole landmarks extracted from 3-D lidar data. While the former two use small proprietary datasets – a fact that makes a direct comparison infeasible – Sefati et al. evaluate their method on sequence number 9 of the



(a) Registration via the original NCLT ground truth.

(b) Refined registration.

Fig. 2: The same set of point clouds taken from a short sequence of an NCLT session, registered using different ground-truth robot poses. The colors encode the point height above ground: Blue represents the ground plane, whereas green, yellow, and red indicate increasing height. The left image shows the result of the registration based on the original NCLT ground truth poses, which we use throughout our experiments. To illustrate the inaccuracy of the original ground truth, the right-hand side image presents a refined registration that we generated via pose-graph optimization. While the original ground truth leads to a blurry point cloud, the refined version significantly improves point alignment and results in crisp details. The mean positional error between both ground truth versions is approximately 0.25 m on average, which leads us to believe that the original NCLT ground truth is off by a similar amount. This fact impedes the generation of an accurate reference pole map and negatively affects our localization results.

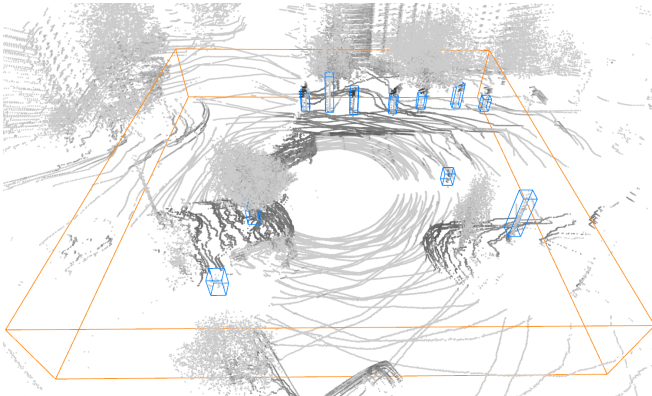


Fig. 3: Exemplary pole extraction result for a point cloud from the NCLT dataset. The gray values of the points correlate with the intensity values returned by the lidar sensor. The orange wireframe represents the boundaries of the local map, while the blue wireframes represent the extracted poles. The pole extractor is triggered by different kinds of pole-shaped objects like traffic signs, street lamps, and tree trunks.

publicly available KITTI dataset [29]. This sequence is a short recording of 46 s along a simple L-shaped trajectory. Trajectories in KITTI have hardly any overlap, which is why Sefati et al. use the sequence data for both mapping and localization. Consequently, their results have limited significance as to real localization performance: They could theoretically localize the vehicle based on dynamic landmarks only, and they would still obtain accurate results with

respect to their map, although it is extremely unlikely that they will encounter the same constellation of dynamic objects ever again. The same is true for Weng et al., who also use a single trajectory of 3.5 km for mapping and localization. Nevertheless, we repeat Sefati et al.’s experiment with the localization system proposed in this paper and compare accuracies in table II. This time, we set the grid spacing for the pole extractor to 0.1 m, because the quality of the ground-truth robot poses is higher than in NCLT. Furthermore, we adjust the parameters of our localizer to match the values Sefati et al. apparently used – 2000 particles, 3 m initial positioning variation, $\pm 5^\circ$ heading variation – and average our results over 50 experimental runs. As shown in table II, our localization system outperforms the reference method by reducing the RMSEs in position and heading by 54% and 69%, respectively. For qualitative analysis, table II also includes the results Kümmerle et al. and Weng et al. obtained after processing their respective proprietary datasets.

V. CONCLUSION AND FUTURE WORK

We presented a complete landmark-based 2-D localization system that relies on poles extracted from 3-D lidar data, that is able to perform long-term localization reliably, and that outperforms current state-of-the-art approaches in terms of accuracy. The implementation is publicly available [1].

For the future, we have two major extensions in mind. First, we plan to fuse the separated mapping and localization modules into a single SLAM module. Second, we would like to explore pole-based localization in different sensor modalities.

Approach	Δ_{pos} [m]	RMSE _{pos} [m]	Δ_{lat} [m]	σ_{lat} [m]	Δ_{lon} [m]	σ_{lon} [m]	Δ_{ang} [°]	σ_{ang} [°]	RMSE _{ang} [°]
Kümmerle et al. [5]	0.12	—	0.07	—	0.08	—	0.33	—	—
Weng et al. [6]	—	—	—	0.082	—	0.164	—	0.329	—
Sefati et al. [8]	—	0.24	—	—	—	—	—	—	0.68
Ours	0.096	0.111	0.061	0.075	0.060	0.067	0.133	0.188	0.214

TABLE II: Comparison of the accuracies of Sefati et al.’s method and the proposed localization approach on the KITTI dataset. The results of Weng et al. and Kümmerle et al. are not directly comparable and are stated for qualitative analysis only.

VI. ACKNOWLEDGEMENTS

We thank Arash Ushani for his kind support with the NCLT dataset.

REFERENCES

- [1] A. Schaefer and D. Büscher. Long-term urban vehicle localization using pole landmarks extracted from 3-D lidar scans. [Online]. Available: <https://github.com/acschaefer/polex>
- [2] M. Modsching, R. Kramer, and K. ten Hagen, “Field trial on GPS accuracy in a medium size city: the influence of built-up,” in *3rd Workshop on Positioning, Navigation and Communication*, vol. 2006, 2006, pp. 209–218.
- [3] N. Carlevaris-Bianco, A. K. Ushani, and R. M. Eustice, “University of Michigan north campus long-term vision and lidar dataset,” *International Journal of Robotics Research*, vol. 35, no. 9, pp. 1023–1035, 2015.
- [4] J. Levinson and S. Thrun, “Robust vehicle localization in urban environments using probabilistic maps,” in *2010 IEEE International Conference on Robotics and Automation*, May 2010, pp. 4372–4378.
- [5] J. Kümmerle, M. Sons, F. Poggendorf, T. Kuehner, M. Lauer, and C. Stiller, “Accurate and efficient self-localization on roads using basic geometric primitives,” in *2019 IEEE International Conference on Robotics and Automation*, May 2019.
- [6] L. Weng, M. Yang, L. Guo, B. Wang, and C. Wang, “Pole-based real-time localization for autonomous driving in congested urban scenarios,” in *2018 IEEE International Conference on Real-time Computing and Robotics*, August 2018, pp. 96–101.
- [7] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [8] M. Sefati, M. Daum, B. Sondermann, K. D. Kreisköther, and A. Kampker, “Improving vehicle localization using semantic and pole-like landmarks,” in *2017 IEEE Intelligent Vehicles Symposium*, June 2017, pp. 13–19.
- [9] R. Spangenberg, D. Goehring, and R. Rojas, “Pole-based localization for autonomous vehicles in urban scenarios,” in *2016 IEEE International Conference on Intelligent Robots and Systems*, October 2016, pp. 2161–2166.
- [10] C. Brenner, “Global localization of vehicles using local pole patterns,” in *Pattern Recognition*, J. Denzler, G. Notni, and H. Süße, Eds. Springer Berlin Heidelberg, 2009, pp. 61–70.
- [11] C. Cabo, C. Ordóñez, S. García-Cortés, and J. Martínez-Sánchez, “An algorithm for automatic detection of pole-like street furniture objects from mobile laser scanner point clouds,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 87, pp. 47–56, 01 2014.
- [12] F. Tombari, N. Fioraio, T. Cavallari, S. Salti, A. Petrelli, and L. D. Stefano, “Automatic detection of pole-like structures in 3D urban environments,” in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, September 2014, pp. 4922–4929.
- [13] B. Rodríguez-Cuenca, S. García-Cortés, C. Ordóñez, and M. C. Alonso, “Automatic detection and classification of pole-like objects in urban point cloud data using an anomaly detection algorithm,” *Remote Sensing*, vol. 7, no. 10, pp. 12 680–12 703, 2015.
- [14] Y. Yu, J. Li, H. Guan, C. Wang, and J. Yu, “Semiautomated extraction of street light poles from mobile LiDAR point-clouds,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3, pp. 1374–1386, 2015.
- [15] F. Wu, C. Wen, Y. Guo, J. Wang, Y. Yu, C. Wang, and J. Li, “Rapid localization and extraction of street light poles in mobile lidar point clouds: a supervoxel-based approach,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 2, pp. 292–305, February 2017.
- [16] H. Zheng, F. Tan, and R. Wang, “pole-like object extraction from mobile lidar data,” *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 41, 2016.
- [17] H. Yokoyama, H. Date, S. Kanai, and H. Takeda, “Detection and classification of pole-like objects from mobile laser scanning data of urban environments,” *International Journal of CAD/CAM*, vol. 13, no. 2, pp. 31–40, 2013.
- [18] C. Ordóñez, C. Cabo, and E. Sanz-Ablanedo, “Automatic detection and classification of pole-like objects for urban cartography using mobile laser scanning data,” *Sensors*, vol. 17, no. 7, p. 1465, 2017.
- [19] F. Li, S. Oude Elberink, and G. Vosselman, “Pole-like road furniture detection and decomposition in mobile laser scanning data based on spatial relations,” *Remote Sensing*, vol. 10, no. 4, 2018.
- [20] B. Qin, Z. Chong, T. Bandyopadhyay, M. H. Ang, E. Frazzoli, and D. Rus, “Curb-intersection feature based Monte Carlo localization on urban roads,” in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 2640–2646.
- [21] A. Schindler, “Vehicle self-localization with high-precision digital maps,” in *2013 IEEE Intelligent Vehicles Symposium*, June 2013, pp. 141–146.
- [22] M. Schreiber, C. Knöppel, and U. Franke, “LaneLoc: lane marking based localization using highly accurate maps,” in *2013 IEEE Intelligent Vehicles Symposium*, June 2013, pp. 449–454.
- [23] A. Y. Hata and D. F. Wolf, “Feature detection for vehicle localization in urban environments using a multilayer lidar,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 2, pp. 420–429, February 2016.
- [24] A. Welzel, P. Reisdorf, and G. Wanielik, “Improving urban vehicle localization with traffic sign recognition,” in *2015 IEEE International Conference on Intelligent Transportation Systems*, September 2015, pp. 2728–2732.
- [25] J.-H. Im, S.-H. Im, and G.-I. Jee, “Vertical corner feature based precise vehicle localization using 3D lidar in urban area,” *Sensors*, vol. 16, no. 8, p. 1268, 2016.
- [26] L. Luft, A. Schaefer, T. Schubert, and W. Burgard, “Closed-form full map posteriors for robot localization with lidar sensors,” in *2017 IEEE International Conference on Intelligent Robots and Systems*, September 2017, pp. 6678–6684.
- [27] K. Fukunaga and L. Hostetler, “The estimation of the gradient of a density function, with applications in pattern recognition,” *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 32–40, January 1975.
- [28] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. The MIT Press, 2005.
- [29] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: the KITTI dataset,” *International Journal of Robotics Research*, 2013.