Abstract—Accurate and reliable localization is a prerequisite for autonomous driving. Methods based on sparse landmarks, such as pole-like structures, have been widely studied because of their lower requirements for computing and storage. However, the number of landmarks of a single type is not always sufficient for reliable positioning. We propose a localization method using three different types of features in urban environments. The features we choose are poles, corners and walls which are persistent over time and can be reliably detected with LiDAR sensors. A pattern matching method for data association is introduced. Instead of using a filtering method, we adopt the graph-based optimization method to solve the pose estimation problem. Experiments conducted on two test roads show that the proposed method can provide accurate and reliable localization results in urban environments.

I. INTRODUCTION

Autonomous vehicles require accurate and reliable localization especially driving in urban environments. Costly GNSS/INS navigation systems are limited by occlusions of the satellite signals, atmosphere changes, and multipath reflections. They cannot always guarantee accurate positioning in urban environments. In order to achieve this, a variety of alternative precise localization methods have been studied, involving sensors such as LiDARs, video cameras, radars, IMUs, or a combination thereof. In particular, multi-beam LiDARs provide fast and accurate measurements of the environment in form of three-dimensional point clouds. They are commonly used for ultrareliable localization and mapping in the field of autonomous driving.

In this paper, we present a localization algorithm based on sparse landmark features obtained with a 3D LiDAR sensor. LiDAR-based features vary in number, spatial distribution, and semantic level, depending on how they are selected. For example, a relatively large number of low-level geometric features can be chosen according to the local curvature or the gradients of surfaces. The detection of 3D objects, such as planes and poles, produces a sparser distribution of high-level features. Such 3D objects have the advantages of being sensor independent. Moreover, these kinds of features are convenient for manual checking of the detections and the correctness of the data association. They have lower storage and calculation requirements. Such features can be provided as part of the HD map of a city from third-party organizations or companies. Pole-like structures are one of the most frequently used features for localization of autonomous vehicles as they are stable across seasons and can be detected easily with a LiDAR system.

II. RELATED WORK

Several industrial companies and research institutes have collaborated to conduct automated and connected driving tests in Berlin. In a residential area for about one kilometer long in the test field, only a few pole-like objects such as streetlamps and trees can be found. They are not dense enough for reliable data association. This situation is common in residential areas outside of main roads. Therefore, we enrich the set of landmarks by adding fragments of walls and corners. Compared to other objects such as curbs, which are often occluded by parking cars, walls can be more reliably detected. The combination of walls, borders of windows and doors, can be detected as corners, which increases the total number of features at our disposal.

Regarding vehicle localization, filtering algorithms such as particle filters, or EKF, have been widely used for real time results. Particle filtering has disadvantages such as low efficiency and particle degeneracy. EKF relies on initial estimates for its linearization expansion, it has a relatively larger error when solving actual inherently nonlinear problems. We adopt the optimization method for pose estimation. Recent studies have explored factor graph-based methods for pose estimation and compared it to filtering methods [1].

The main contributions of this work are the following: (a) We provide a method for accurate and reliable localization for automated driving in urban environments. (b) The features we choose are three types of landmarks, which largely increase the number of features compared to solely pole-based methods. They can localize a vehicle in a residential area with few poles. (c) The results of our experiments demonstrate good repeatability and high accuracy.

All Authors are with the Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin, 14195, Germany
took advantages of reflection density information from laser scanners to detect various markings on the road surface, in conjunction with curb maps for vehicle positioning. In [7], the authors also used lane markings and curbs for localization, but the features were detected with a stereo camera. This study demonstrated accurate positioning result on rural roads.

Landmark-based localization and SLAM algorithms usually use the nearest neighbor method for data association, and it can be problematic when the initial position is inaccurate. Mismatches can also happen when the feature detection is noisy. One solution to this issue is to use the relations between local features for pattern matching. Previous research [8], [9] used the point pattern for data matching. Study [4] introduced a matching method using point and line features. Drawing from the related work, for this research, we use three different types of features, and therefore a new matching method is required.

Although different types of features have been studied. For data fusion and pose estimation, particle filters and EKF are most commonly used. The possibilities of using graph-based optimization for localization were explored in [10] and showed its increased robustness against outliers. In [1], factor graph-based localization using third-party maps for automated driving was recently studied, and it presented higher accuracy compared to particle filter methods.

III. SYSTEM OVERVIEW

In this section, we introduce the main parts of our proposed localization method in urban environments, including feature detection methods, a priori map creation, data association and pose estimation methods.

A. Feature Detection

The feature detection algorithm starts from generating range image from LiDAR measurements. The measurement points from the LiDAR sensor arrive in separate packages. We first need to eliminate the motion distortion while the car is moving. The wheel odometer provides movement information with a frequency of 100 Hz. It can be used to align LiDAR measured points. When a 360-degree scan is completed, the measurements are combined as a range image. Fig. 1 shows part of the range image after the motion distortion correction. The feature detection process can be outlined in two steps. First, the image is scanned line by line to find potential feature segments of pole, wall and corner. Feature segments of each type are stored in a segment buffer for follow-up processing. The second step is to merge the buffered segments into features.
plane, they basically form a corner. Fig. 3 shows the detected features: green cylinders are pole-like features, yellow planes are wall segments, and magenta columns are corners.

B. Feature Map

The localization algorithm needs a high precision \textit{a priori} map for feature matching. The quality of this map directly affects the accuracy of the localization. Our test vehicle is equipped with an Applanix DGPS/INS navigation system, which can provide global position for localization and mapping. However, it can not always provide accurate data to create consistent maps. We adopted a SLAM method to produce our feature-based \textit{a priori} map. The off-line batch solution globally minimizes motion and observation errors and creates a self-consistent map. We implemented a graph SLAM with loop closure aided by the Applanix navigation system as it provides highly accurate angular and linear velocity. Previous research [11] presented a GNSS/INS aid SLAM method to map trees, which gives inspiration for our work. We adopted the graph-based optimization framework g2o [12] in this study. For feature extraction we applied the method introduced above. Fig. 4 and Fig. 5 show the created feature map of two test tracks from previously recorded data. Our experiments are conducted on these two tracks and the results will be explained in detail in the next section.

C. Data Association

Extensive research has been conducted to explore the problem of matching two sets of 2D objects under arbitrary translations, rotations, and scale changes. It is widely used in fingerprint recognition [13], object recognition [14], image matching, scene recognition, and other areas. We use a pattern matching method to find the association between a set of detected features and the features in the \textit{a priori} map. Different from a common affine invariant matching in computer vision area, the features detected by LiDAR have no scale changes, and the point features are of two different types (pole and corner). For our application scenario, we propose a new method for map matching.

Before each feature registration, we first update a local grid feature map. The map data within the sensor’s detection range are encoded as a grid map as shown in Fig. 6. Through indexes, we can find out whether there is a certain type of feature that exists in a grid cell with a time complexity of \textit{O}(1).

For each feature detection, detected features are denoted as three sets: \(S_p, S_c, S_w\), to represent the poles, corners and walls. Sets \(S_p, S_c, S_w\) contain three types of map features in the local area within the sensor range. Further we define the tuple sets \(t_p = \{(p_i, p_j, d_{ij}) | p_i, p_j \in S_p, i \neq j\}\) and \(t_c = \{(p_i, p_j, d_{ij}) | p_i, p_j \in S_c, i \neq j\}\) where \(d_{ij}\) is the distance between \(p_i\) and \(p_j\). Similarly, we define tuple sets \(T_p, T_c, T_{pc}\) for the local map data. Algorithm 1 computes the matched points of detected features and the map elements. The transformation between these two sets of features are also calculated.

D. Graph Optimization for Pose Estimation

Next, we introduce how to use a graph-based optimization to estimate vehicle poses with 3-DOF. Fig. 7 illustrates the problem in form of a graph. Very similar to the SLAM problem, we denote the state as \(x = [x_1, x_2, \ldots, x_K, m_1, m_2, \ldots, m_M]^T\), where \(K\) is the size of poses

![Fig. 6. Left: part of the feature map, it is encoded to local grid map (right) for fast retrieve.](image-url)
including in each optimization, M is the total number of features that can be observed within these poses. The maximum a posteriori $x_{MAP}$ can be written as:

$$x_{MAP} = \arg \max_x p(x|u,y).$$  

(1)

Here $u$ is the motion of the vehicle between two adjacent poses, we use the data from the wheel odometer in this study, and $y$ is the measurement of all observed features within $K$ poses. Applying Bayes theorem to (1), we get:

$$x_{MAP} = \arg \max_x p(x|u,y) = \arg \max_x \frac{p(y|x)p(x|u)}{p(y|u)}$$

(2)

The motion function and measurement function are:

$$x_k = f(x_{k-1}, u_k, w_k)(k = 1, ..., K)$$

(3)

$$y_{kj} = h(x_k, m_j, v_{kj})(k = 1, ..., K; j = 1, ..., M).$$

(4)

Assume the motion noises $w_k$ and $v_{kj}$ to be independent and additive. $R_k$ and $Q_{kj}$ are the covariances of $w_k$ and $v_{kj}$ respectively. Now the motion and measurement error can be written as follows:

$$e_{f,k} = x_k - f(x_{k-1}, u_k)$$

(5)

$$e_{h,kj} = y_{kj} - h(x_k, m_j).$$

(6)

Take the negative logarithm to $p(y|x)p(x|u)$ in (2)

$$-\ln(p(y|x)p(x|u)) = -\sum_{k=1}^{K} \ln p(y_k|x_k) - \sum_{k=1}^{K} \ln p(x_k|x_{k-1}, u_k).$$

(7)

After substitution of (5), (6) to (7), removing the constant terms and normalizing it, we can get:

$$J(x) = \sum_{k} e_{f,k}^T R_k^{-1} e_{f,k} + \sum_{k} \sum_{j} e_{h,kj}^T Q_{kj}^{-1} e_{h,kj}.$$  

(8)

Then we consider the constraints from the a priori map. As shown in Fig. 7, $p_j$ represents a priori information from the prebuilt feature map. Assume the covariance of the a priori map is known as $S_k$. We can now simply calculate the map error as

$$e_{m,j} = m_j - p_j.$$  

(9)

For point features, $e_{m,j}$ is the position difference of two points, and for wall features $e_{m,j} = (\Delta d_j, \Delta \theta_j)$. By adding the map error to (8), we can get the final objective function:

$$J(x) = \sum_{k} e_{f,k}^T R_k^{-1} e_{f,k} + \sum_{k} \sum_{j} e_{h,kj}^T Q_{kj}^{-1} e_{h,kj} + \sum_{k} \sum_{j} e_{m,j}^T S_j^{-1} e_{m,j}.$$  

(10)

Our goal is to minimize $J(x)$. We use the g2o [12] library to build graphs like Fig. 7 and use the Levenberg-Marquardt algorithm to iteratively solve the problem. The number of poses taken into optimization is limited by the size of the sliding window. In SLAM problems, to ensure computation efficiency, a marginalization technique is usually used to limit the number of states involved in the computation. Here we directly discard the previous states. The reason is that our a priori map built from the offline SLAM approach is self-consistent, we give high confidence in it. As a result, it provides strong constraints on the current states. If a large size for the sliding window is selected, the previous constraints have little effect on the current states. And this can also avoid the accumulation of linearization errors caused by marginalization.

IV. RESULTS

We ran our method on an autonomous test vehicle and collected datasets. The vehicle is equipped with a Velodyne HDL-64E LiDAR scanner, an Applanix POS-LV 510 navigation system, and other sensors such as cameras, radar and other laser scanners. The algorithm was run on a laptop with an Intel Core i7-7700HQ CPU. A GPS position was required at the beginning to initialize the algorithm. The method provides 2D position and heading information at a frequency of 100Hz.

We use two test roads. Test Road 1 is “Thielallee”, a street near our research institute. The length of one lap on this road is about 1.2 km. There are plenty of trees along the road. Test Road 2 is a test area for connected and automated driving in Berlin. It contains around 5.6 km main roads with complex traffic conditions. Fig. 4 and Fig. 5 show the feature map of
TABLE I
CORRECT MATCHING RATE AND AVERAGE CALCULATION TIME OF THE PROPOSED PATTERN MATCHING METHOD.

<table>
<thead>
<tr>
<th>Road</th>
<th>Correct Matching</th>
<th>Average Matching Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Road 1</td>
<td>99.53 %</td>
<td>0.27 ms</td>
</tr>
<tr>
<td>Test Road 2</td>
<td>98.65 %</td>
<td>0.45 ms</td>
</tr>
</tbody>
</table>

Three experiments were conducted. First, we tested our map matching method. We obtain feature detection results from LiDAR scanner every 0.1 second on average. The feature matching results on the Test Road 2 is shown in Fig. 8. The average number of successfully matched poles in this area was 13.9 at each time, and this number increased to 21.4 if all three types of features were used. We can find that on the left side of the test road, the number of successfully associated poles is under 5 for most parts of this region. It is significantly improved if corners and walls are added.

Table I shows the matching success rate on two test roads. A successful matching means at least half (6 at minimum) of the detected features are associated with feature map, and these associations lead to correct transformation result. The correct matching rate are 99.52% and 98.65% on the two test roads, respectively. Failed match is usually due to the number of detected features is insufficient to match with the feature map reliably. In the case of a failed match, the corresponding pose has only a constraint from wheel odometer. In terms of running time, the ClusterMatch algorithm has a very high efficiency and is sufficient to support real-time localization.

Compared with the commonly used Nearest Neighbor matching method, the ClusterMatch has several advantages. First, it is not sensitive to the accuracy of the initial position. In fact, it can perform global matching without knowing the initial position. But in practical, the detected features are only matched with a local sub-map in order to achieve a high computation efficiency. The size of the sub-map is the effective feature detection range with an extension, which is 60 meters around current position in total. Therefore, the initialization of the proposed method requires only a rough position, and an ordinary GPS is enough. Secondly, the Nearest Neighbor method requires high accurate and reliable a priori information of current pose before conducting the map matching. Otherwise, the correct data association may not be obtained. ClusterMatch performs feature matching within the scope of local sub-maps, and data association does not depend on the accuracy of pose prediction.

In the second experiment, we compared the feature localization results with the trajectory reported by the navigation system. With the RTK (Real-Time Kinematic) technique, it can provide a comparably precise positioning information. However, analysis of the results in the test roads showed that only when the vehicle is stationary, the error may fall below 10 cm. A larger error occurred during movement, and sometimes it showed an unpredictable drift that can exceed 1 meter. This is a situation that our solution needs to avoid. Therefore, we conducted a repeatability test to evaluate the lateral accuracy. The test method was to let the human driver repeatedly drive on the short test road and try to keep the vehicle in the center of the lane. The repeatability can be estimated from the recorded positioning trajectories. As shown in Fig. 9, the localization method proposed shows better consistency. After removing the U-turn area where it is difficult to maintain a consistent driving trajectory, the average lateral error is around 15 cm. Considering the error introduced by the human driver, this result is acceptable. The error of positioning results reported by the Applanix system are significantly larger.

On the third experiment we evaluated the accuracy of our localization algorithm. As shown in the above experiment, the navigation system has an unpredictable drift error, so we did not use it as the ground truth. We ran the offline SLAM algorithm introduced earlier on the recorded test dataset to get optimized poses with timestamps as ground truth. Then we could compare our localization performance against it. Fig. 10 shows the distance error over time. Compared with the navigation system, the proposed feature localization method has a smaller distance error most of the time. And its results show a smaller variance.

Table II outlines the experiment results of the average distance error, RMSE and maximum distance error of the two methods. On Test Road 1, the average distance error and RMSE of the proposed method is 10.2 cm and 11.8 cm, respectively. The maximum distance error is 32.0 cm. On Test Road 2, the error became larger. It increased to 15.2 cm and 17.1 cm. The maximum distance error is 37.2 cm on this road. The Applanix system shows a relatively larger error on both test roads, especially on Test Road 2. The reason could be that Test Road 2 is in a residential area with plenty of buildings along the road, which may affect GPS signals. It can be seen that the proposed method presents accurate localization results, showing advantages over the positions reported by the Applanix system.

V. CONCLUSION

We introduced a LiDAR based localization method for autonomous vehicles in the urban environment. The features
Fig. 9. Trajectory plot of feature localization results and Applanix reported positions.

TABLE II

<table>
<thead>
<tr>
<th>Method</th>
<th>Test Road 1</th>
<th>Test Road 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average Distance Error</td>
<td>RMSE</td>
</tr>
<tr>
<td>Applanix</td>
<td>0.162</td>
<td>0.211</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.102</td>
<td>0.118</td>
</tr>
</tbody>
</table>

Fig. 10. Distance error over time, on Test Road 1 (top) and Test Road 2 (bottom).

we used were pole-like structures, corners and walls. We also outlined the pattern-matching-based data association method and the optimization method for pose estimation. Experiments were conducted in practice and on datasets. The proposed method showed its capability of accurate localization in residential area where only a few poles exist. In both practical and dataset tests, the proposed method presented better repeatability and accuracy compared to a high-cost navigation system.

REFERENCES


