# Vehicle Localization on Semantic Map by Combining Visual and Distance Measurements

Zhihuang Zhang, Changyao Huang, and Liang Li(✉)

Tsinghua University, Beijing 100084, China
linagl@tsinghua.edu.cn

**Abstract.** Nowadays, a precise and robust localization system is essential for an autonomous vehicle to operate safely and effectively. Currently, individual sensor still can not reach a satisfying level in both accuracy and robustness. It is a feasible solution to build a localization system by fusing multi-type sensors and incorporating map information. This paper proposes a real-time multi-sensor fusion method based on the particle filter framework. It utilizes the onboard IMU, camera, lidar, and digital map information. The main novelty and contribution of this paper lie in combining visual and distance measurements into semantic features to correct the states. From a mass production perspective, the proposed system is satisfactory in terms of accuracy and map volume. The proposed method is quantitatively evaluated in real complex scenarios. The experimental results show the effectiveness and accuracy of the proposed system.

**Keywords:** Vehicle localization · Semantic map · Map matching · Sensor fusion

## 1 Introduction

Vehicle localization is crucial to autonomous driving. The localization system measures the position, velocity and pose of an autonomous vehicle. Precise localization enables the perception system to better perceive the driving environment. Planning module also requires correct states to generate traveling paths and driving controls [4]. Therefore, the autonomous driving system needs a localization system with relatively high positioning accuracy and robustness. The positioning estimation is conducted by utilizing information of onboard sensors. Typical sensors for positioning are the global navigation satellite system (GNSS) receivers and the inertial measurement units (IMU). However, low grade sensors provide inadequate precision since they are affected by larger noises and biases [2,10,13].

Fortunately, the localization system is possible to benefit from a pre-built digital map with the map-matching process [15]. The map matching process extracts environmental features and associates the extracted features with map features to compute optimal position estimates. The precision relies on both


W. Fu et al. (Eds.): ICAUS 2022, LNEE 1010, pp. 1583–1590, 2023.
https://doi.org/10.1007/978-981-99-0479-2_146

sensors and map performance. With the advancement of neural network-based perception technology, the semantic feature-based methods have dominated the camera-based localization systems [7,9,11]. The semantic features are more robust against environmental changes and bring lower capacity requirements of the digital map [14]. However, the major drawback of visual measurement is its inability to directly obtain distance measurements of features, which explains why visual positioning systems tend to have large positioning errors. In contrast, lidar measures environmental features in three dimensions with high precision. Previous researches match the current scans to dense point cloud maps by the iterative closest point (ICP) method or normal distributions transform (NDT) method to calculate position estimates [8,12]. Besides, learning-based methods have also shown their talents in terms of accuracy [5]. Although the point cloud-based method provides higher accuracy, their gap lies in the dense and heavy maps, which limits the application for mass production. Consequently, combining visual and distance measurements to localize position on semantic maps is promising for improving accuracy while retaining lightweight properties.

Besides, a single sensor or map-matching module could not meet the accuracy and robust requirements of the autonomous driving system. Therefore, researchers take advantage of multi-sensor fusion to obtain better solutions. The central idea of multi-sensor fusion is to make full use of measurement information and determine a solution with the least error variance. The Kalman filtering-based methods have been extensively researched owing to their efficiency [1,3,6]. However, during the map matching processes, some of the features cannot be directly used to update the states. Therefore we turn to the particle filter as the sensor fusion framework.

In this paper, we propose a real-time method to localize the vehicle position on semantic maps. A particle filter is used to fusion onboard sensors. The main contribution lies in combining visual and distance measurements into semantic features to correct the particle states. The proposed method is quantitatively evaluated in real complex scenarios. The experimental results indicate that the localization system delivers accurate position and pose accuracy, which meets the requirements of autonomous vehicles.

## 2 Method

### 2.1 System Overview

Considering an autonomous vehicle equipped with onboard sensors. The sensors consist of GNSS receiver, low-cost IMU, wheel speed sensor, front camera, and lidar. The goal of the localization system is to estimate the vehicle position and pose on a digital map. Figure 1 shows an example of the digital semantic map. The map consists of poles, lane line. These features contain the information of the absolute location, which can be used to restore the vehicle position.

A particle filter is built to fuse these signals. The wheel speed signal and the IMU signal are combined to update the state of the particles. Since the low-cost GNSS receiver could not provide adequate accuracy for position correction, we turn to obtaining position measurement through map-matching.

**Fig. 1.** An scenario of HD digital map, the map describes the road topology and geometry information.

### 2.2 Visual and Distance Measurements Combination

In this section, the details of the visual and distance measurements combination are introduced. Considering an image $\mathcal{I}_k$ obtained from the camera at time $k$, the goal is to associate the visual semantic information to the points of a lidar scan $\mathcal{L}_k$. Note that laser scans may not be captured at time $k$, motion compensation is required before starting the association.

For every point $p_i^L$ in the laser scan, we project the point into the image plane as follows:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = p_i^I = K T_I^L p_i^L \tag{1}$$

where $K$ is the camera intrinsics matrix, $T_I^L$ is the transform matrix from lidar coordinate system to camera coordinate system.

As shown in Fig. 2, the image is processed by a features detection network, which extract lane line and pole features. With the extracted features, the transformed points can be labeled with a specific feature type. A pixel threshold is set to filter out points that are excluded from the target feature.

Nevertheless, the association in the image plane may gather points that do not belong to the same feature. Therefore in the next stage, the points are further filtered in the lidar frame. For the points belonging to the lane line, the points are filtered according to their horizontal height. For the points belonging to the pole, we perform clustering in the vertical direction to ensure that these points belong to a vertical line. After the filtering, the extracted features $Z$ can be used to corrected the states in particle filter.

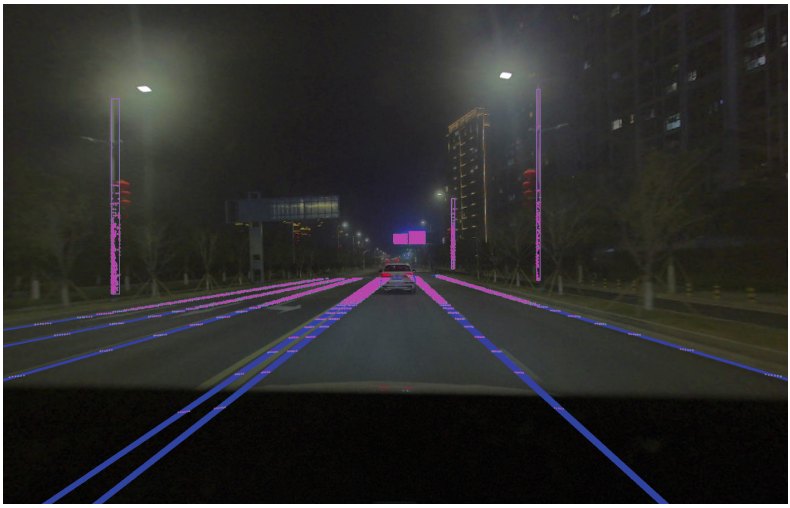

**Fig. 2.** The extracted semantic features are used to label transformed lidar points, the outlier points are dropped out.

$$Z = \left\{ z_i^{pole}, z_j^{line} \right\}_{i=1:N_{pole}, j=1:N_{line}} \tag{2}$$

### 2.3 Particle Based Sensor Fusion

This section focus on the particle filter based sensor fusion. Since the horizontal position accuracy is more concerned, the state vector in our particle filter is defined as $X_k^i = [x_k^i, y_k^i, \psi_k]^T$, where $x_k^i$, $y_k^i$ and $\phi_k^i$ are the $i$th particle's 2D position and heading angle in the east-north-up (ENU) coordinate system. We adopt the kinematic model to describe the motion of the vehicle as follows:

$$X_{k+1}^i = \begin{bmatrix} \tilde{x}_{k+1}^i \\ \tilde{y}_{k+1}^i \\ \tilde{\psi}_{k+1}^i \end{bmatrix} = \begin{bmatrix} \tilde{x}_k^i + \tilde{v}_k \Delta t \cos\left(\tilde{\psi}_k^i + \frac{\tilde{\omega}_k \Delta t}{2}\right) + \delta_x^i \\ \tilde{y}_k^i + \tilde{v}_k \Delta t \sin\left(\tilde{\psi}_k^i + \frac{\tilde{\omega}_k \Delta t}{2}\right) + \delta_y^i \\ \tilde{\psi}_k^i + \tilde{\omega}_k \Delta t + \delta_\psi^i \end{bmatrix} \tag{3}$$

where $\tilde{v}_k$ is the vehicle speed measured by the wheel speed sensor. $\tilde{\omega}_k$ represents the $z$-axis angular velocity acquired from IMU. $\delta_x^i$, $\delta_y^i$, $\delta_\phi^i$ stand for the increment of uncertainty during the prediction stage.

The filter requires an initialization when the localization system boots from a cold start. For each particle, their states are randomly generated following the Gaussian distributions whose mean and variance depend on the GNSS measurement.

In the update phase, the particle weights are sequentially updated by the measurements. The updating equation is:

$$\tilde{\omega}_{k+1}^{i} = \omega_{k}^{i} \cdot \omega_{k,pole}^{i} \cdot \omega_{k,lane}^{i} \cdot \omega_{k,heading}^{i} \tag{4}$$

where $\omega_{k,pole}^{i}$, $\omega_{k,lane}^{i}$ and $\omega_{k,heading}^{i}$ represent the weight offerd by different type of features.

$$w_{k,pole}^{i} = \frac{1}{2\pi\sqrt{\det\left(R_{pole}\right)}} \exp\left[-\frac{1}{2}\left(\hat{r}_{k}^{i} - r_{k}\right)^{\mathrm{T}} R_{k,pole}^{-1}\left(\hat{r}_{k}^{i} - r_{k}\right)\right] \tag{5}$$

where $R_{pole}$ represents the $2 \times 2$ uncertainty matrix of pole detection. $\hat{r}_{k}^{i}$ is the relative position between the $i$th particle and pole. $r_k$ is the pole position from by the visual and lidar combination module. The pole position is calculated by the mean of $z^{pole}$. We use the nearest neighbor matching in this process.

When processing the lane line, the relative distance between particle and lane line is extracted as $\hat{d}_{k}^{i}$. The percepted relative distance is $d_k$, and the uncertainty is denoted as $\sigma_{lane}$. The heading angle update follows the same pattern, note that the relative heading angle is calculated by the curvature of lane line.

$$w_{k,lane}^{i} = \frac{1}{\sqrt{2\pi\sigma_{lane}^{2}}} \cdot \exp\left[-\frac{1}{2}\left(\hat{d}_{k}^{i} - d_{k}\right)^{2} / \sigma_{lane}^{2}\right] \tag{6}$$

$$w_{k,heading}^{i} = \frac{1}{\sqrt{2\pi\sigma_{\psi}^{2}}} \cdot \exp\left[-\frac{1}{2}\left(\hat{\psi}_{k}^{i} - \psi_{k,\psi}\right)^{2} / \sigma_{\psi}^{2}\right] \tag{7}$$

Reweighting is needed in each step of updating to make sure the sum of the particle weights equals to 1. For each particle, the reweighting procedure is:

$$\hat{\omega}_{k}^{i} = \frac{\tilde{\omega}_{k}^{i}}{\sum_{j=1}^{N} \tilde{\omega}_{k}^{i}} \tag{8}$$

Resampling is important in the particle filter as it prevents the situation where the weights concentrate on only a few particles. A random resampling method is utilized when the inequation (9) is satisfied. Where N is the number of the particles.

$$\frac{1}{\sum_{i=1}^{n}\left(\omega_{i}^{2}\right)} < 0.5\,N \tag{9}$$

The estimation equations of position and heading angle are as below:

$$\hat{X}_{k} = \begin{bmatrix} \hat{x}_{k} \\ \hat{y}_{k} \\ \hat{\psi}_{k} \end{bmatrix} = \sum_{i=1}^{N} \hat{\omega}_{k}^{i} \begin{bmatrix} \hat{x}_{k}^{i} \\ \hat{y}_{k}^{i} \\ \hat{\psi}_{k}^{i} \end{bmatrix} \tag{10}$$

Note that the shape of the particle distribution has to be inspected before the output, and we consider the filter as diverging when a bimodal distribution with large distances exists and needs to be re-initialized.

## 3 Experiment Results and Discussions

In this section, we evaluate the performance of the proposed localization system. 4 sequences of driving routes were established and evaluated. The test vehicle is driven following the routes while the onboard computer collects the sensor data and conducts the fusion algorithm. The frequencies of the wheel speed sensor, IMU, camera, and lidar 50 Hz, 100 Hz, 20 Hz, 10 Hz, respectively. Besides we use a high-end GNSS/INS device to generate reference trajectories.

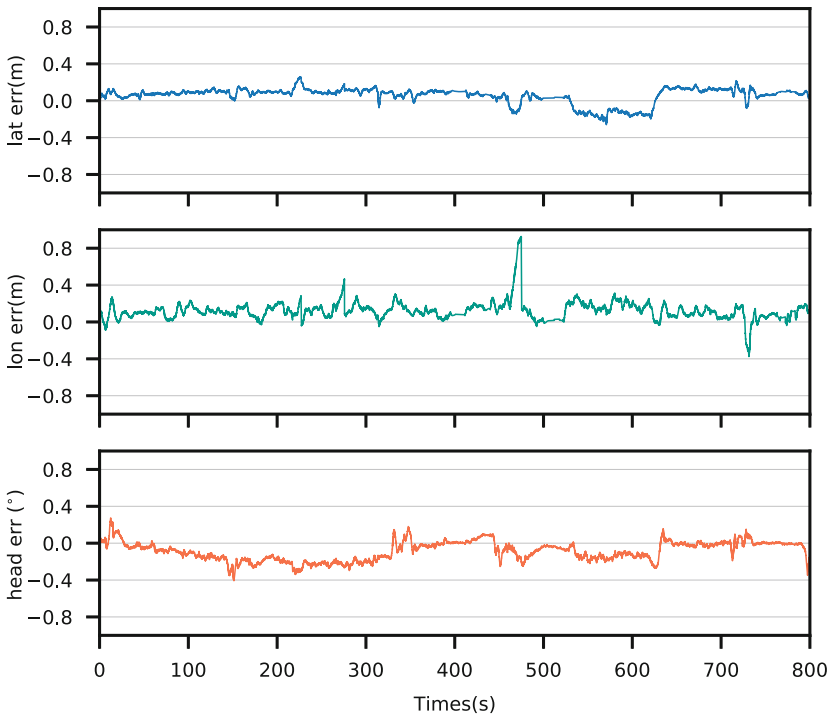


**Fig. 3.** The error curves under the body frames.

### 3.1 Qualitative Analysis

From the perspective of the planning module, the position errors along lateral and longitudinal are more concerned. Therefore, the position errors in the ENU frame are transformed into the body frame. Figure 3 shows the error curves during a time interval. Blue, green, and orange curves represent the lateral position error, the longitudinal position error, and the heading angle error, respectively. It can be seen that the curves have small fluctuations, which indicates the proposed system presents smooth position and heading angle accuracies.

**Table 1.** Localization mean absolute error of different methods.

| | Lat [95%] (m) | Lon [95%] (m) | Head [95%] (°) |
|---|---|---|---|
| Sequence 1 | 0.054 [0.172] | 0.172 [0.411] | 0.142 [0.337] |
| Sequence 2 | 0.061 [0.210] | 0.195 [0.509] | 0.174 [0.372] |
| Sequence 3 | 0.063 [0.198] | 0.189 [0.483] | 0.138 [0.294] |
| Sequence 4 | 0.057 [0.239] | 0.155 [0.395] | 0.135 [0.303] |
| Total | 0.059 [0.218] | 0.179 [0.480] | 0.144 [0.328] |

To further evaluate the accuracy quantitatively, the mean absolute errors are calculated. As shown in Table 1, the proposed method delivers larger longitudinal errors than lateral errors on average. The reason comes from the distribution of the features used to correct the states. In the update stage, we use the pole and the lane line features to update the particle weights. The poles provide lateral and longitudinal restraints, while the lane lines provide only lateral restraints. Since the lane lines are present in most of the scenes, the lateral positions can be continuously corrected. Fortunately, the downstream module demands lower longitudinal accuracy. It can be concluded that the proposed localization system meets the requirements of the autonomous driving system.

### 3.2 Limitations and Future Works

As discussed above, the proposed system works well in most scenes. However, there are some occasions that may lead to system failure. One example is when the vehicle travels through a long interaction without lane lines and poles. In this case, the proposed system can only depend on the prediction process to predict future states. It would not last long as the measurements of low-cost IMU and wheel speed are usually superimposed with high-frequency noises. Adding extensive environmental features to the map to provide more persistent constraints is a way to solve this problem.

## 4 Conclusion

In this paper, we propose a precise vehicle localization system that fuses onboard IMU, camera, lidar, and digital map information. The main novelty and contribution of this paper are combining visual and distance measurements into semantic features to correct the states. Consequently, the position accuracy is improved and the system still retains a lightweight property. From a mass production perspective, the proposed system is adventurous in terms of accuracy and map volume. The system is quantitatively evaluated in complex environments on a real-time platform. Experimental results show that this system offers reliable solutions of the vehicle position in both lateral and longitudinal directions. In future work, we will aim to extend the types of features and hence provide more constraints to improve system performance.

## References


1. Crassidis, J.: Sigma-point Kalman filtering for integrated GPS and inertial navigation. IEEE Trans. Aeros. Electron. Syst. **42**(2), 750–756 (2006)
2. Groves, P.D.: Principles of GNSS, inertial, and multisensor integrated navigation systems. IEEE Aeros. Electron. Syst. Maga. **30**(2), 26–27 (2015)
3. Hartley, R., Ghaffari, M., Eustice, R.M., Grizzle, J.W.: Contact-aided invariant extended Kalman filtering for robot state estimation. Int. J. Rob. Res. **39**(4), 402–430 (2020)
4. Houts, S.E., Cammarata, R., Mills, G., Agarwal, S., Vora, A.: Localization requirements for autonomous vehicles. SAE Int. J. Connect. Autom. Veh. **2**(3), 173–190 (2019)
5. Lu, W., Wan, G., Zhou, Y., Fu, X., Yuan, P., Song, S.: DeepVCP: an end-to-end deep neural network for point cloud registration. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 12–21 (2019)
6. Qin, C., Ye, H., Pranata, C.E., Han, J., Zhang, S., Liu, M.: LINS: a lidar-inertial state estimator for robust and efficient navigation. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 8899–8906 (2020)
7. Suhr, J.K., Jang, J., Min, D., Jung, H.G.: Sensor fusion-based low-cost vehicle localization system for complex urban environments. IEEE Trans. Intell. Transp. Syst. **18**(5), 1078–1086 (2017)
8. Wan, Get al.: Robust and precise vehicle localization based on multi-sensor fusion in diverse city scenes. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 4670–4677 (2018)
9. Wang, H., Xue, C., Zhou, Y., Wen, F., Zhang, H.: Visual semantic localization based on HD map for autonomous vehicles in urban scenarios. In: 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 11255–11261 (2021)
10. Wen, W., Pfeifer, T., Bai, X., Hsu, L.T.: Factor graph optimization for GNSS/INS integration: a comparison with the extended Kalman filter. Navigation **68**(2), 315–331 (2021)
11. Xiao, Z., Yang, D., Wen, T., Jiang, K., Yan, R.: Monocular localization with vector HD map (MLVHM): A low-cost method for commercial IVs. Sensors **20**(7), 1870 (2020)
12. Zhang, J., Singh, S.: LOAM: lidar odometry and mapping in real-time. In: Robotics: Science and Systems X. Robotics: Science and Systems Foundation (2014)
13. Zhang, Z., Zhao, J., Huang, C., Li, L.: Learning end-to-end inertial-wheel odometry for vehicle ego-motion estimation. In: 2021 5th CAA International Conference on Vehicular Control and Intelligence (CVCI), pp. 1–6 (2021)
14. Zhang, Z., Zhao, J., Huang, C., Li, L.: Learning visual semantic map-matching for loosely multi-sensor fusion localization of autonomous vehicles. IEEE Trans. Intell. Veh. **8**, 358–367 (2022)
15. Zheng, S., Wang, J.: High definition map-based vehicle localization for highly automated driving: Geometric analysis. In: 2017 International Conference on Localization and GNSS (ICL-GNSS), pp. 1–8. IEEE, Nottingham (2017)