AUTONOMOUS VEHICLES LOCALISATION BASED ON SEMANTIC MAP MATCHING METHOD

He Huang [1], Dongdong Yu [1], Junxing Yang [1] * , Xun Liu [1]

[1] School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, China - huanghe@bucea.edu.cn, 2108160121004@stu.bucea.edu.cn, yangjunxing@bucea.edu.cn, 2108160120002@stu.bucea.edu.cn

KEY WORDS: Semantic Map, Visual Odometry, Map Matching, Pose Optimization, Relative Localization Accuracy.

ABSTRACT:

In autonomous driving systems, a positioning method that can be used in scenarios with no satellite signals or long signal interruptions is a must. In this paper, we address the problems in map construction methods and map matching methods in scenes without satellite signals or long signal interruptions, and construct a semantic map matching-based localization method to meet the localization requirements by means of monocular vision sensors on the basis of weighing the accuracy and cost of map construction and localization. In this paper, firstly, the method of map construction is studied and a static semantic map construction method based on monocular camera is constructed. Then the map matching localization method is studied, and a semantic map matching based localization method is constructed to align the local map built during localization with the pre-built global semantic map to obtain the current location information. Finally, this paper constructs a method to fuse the visual odometry and map matching localization results, so as to obtain more accurate localization results.

1. INTRODUCTION

In autonomous driving system, accurate real-time acquisition of the vehicle's position and attitude is critical, and accurate positional position is the basis for planning and control decisions. In general, combining GNSS with inertial measurement units and using real-time differential technology can provide accurate, real-time position information for autonomous driving systems. However, in some urban scenarios, the satellite signal is susceptible to interference or occlusion, resulting in discontinuous reception of the satellite signal. The situation can cause the combined navigation system to degrade rapidly and produce large errors, which seriously affects the positioning accuracy and makes the positioning results no longer applicable to the subsequent decision judgment of the autonomous driving.Therefore, a localization method that can be used in scenarios without satellite signals or with prolonged signal interruptions is necessary for autonomous driving systems.

Semantic map-based matching localization, as a popular problem for unmanned collar research in recent years, has accumulated some experience, and a variety of algorithms for map matching localization have been proposed. Feature point method SLAM is currently more mature, such as ORB-SLAM3 (Campos C et al.2020) is able to represent the whole image by selecting feature points in the image, which can work used when the noise is high or the camera movement is fast, and many feature points usually have better robustness to scale change and rotation, and also insensitive to the change of illumination in the environment. VINS-Mono (Tong et al.2018) is a sliding window-based SLAM algorithm that uses monocular cameras and IMU sensors for simultaneous localization and map construction. However, the feature point method needs to extract feature points and calculate descriptors, which will consume more computation time. The direct method SLAM such as LSD-SLAM (Fleet D et al.2014) does not need feature extraction and operates on pixels directly, which can theoretically utilize all the information of the image, but the direct method is based on the assumption of grayscale invariance, which causes it to be very sensitive to changes in ambient illumination and more affected by changes in illumination intensity.

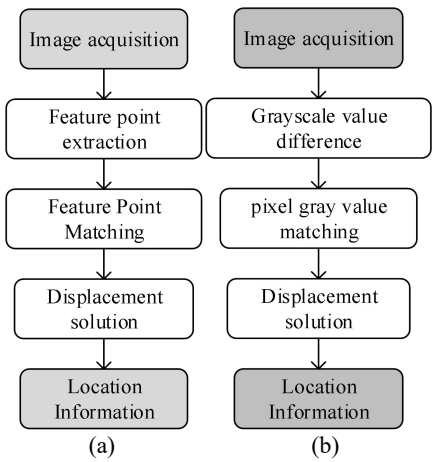

Figure 1. (a)Visual odometry with feature point method
(b) Visual odometry with direct method

Map matching algorithms such as Road-SLAM (Jeong J et al.2017) and RoadMap (Qin T et al.2021) have the following problems: firstly, the contradiction between the fully utilized semantic information and the map updating efficiency, which is difficult to guarantee the effectiveness when the map information is complete and requires high update frequency; then, the current map construction relies too much on GNSS systems to provide real-time positional information, and the positioning method in the satellite signal blocked area needs Finally, when there are unstable dynamic objects in the environment, it will affect the accuracy of positioning and map construction and reduce the quality of final positioning.

* Corresponding author

For semantic recognition and segmentation of road scenes, a real-time high-performance DCNN-based (Dong G et al.2018) semantic segmentation method for urban street scenes, which achieved a good balance between accuracy and speed. a new deep dual-resolution network for real-time semantic segmentation (Hong Y et al. 2019) of road scenes. The method achieves a new optimal balance between accuracy and speed on the Cityscapes dataset and the CamVid dataset, reaching the latest optimal performance. PP-Liteseg (Peng J et al.2022), in which a flexible and lightweight decoder is used to reduce the computational burden. And in order to enhance the feature representation, a unified attention fusion module is added, which uses spatial and channel attention to generate weights, and then fuses the input features with the weights, combined with a simple pyramid pooling module to aggregate the global context with low computational cost.

To address the problems of autonomous driving localization systems in satellite signal blocked areas at the current stage, this paper investigates how to build a lightweight, accurate map containing the necessary localization semantic information based on some currently available localization and map construction techniques, and completes the localization process by map matching based on the built map. In this paper, the semantic map construction algorithm and the map matching based positioning algorithm are implemented with the data of road environment scenes in the park obtained from open source datasets and self-built collection platforms, and experiments are designed to verify the accuracy of map construction and positioning.

Our main contributions are as follows:
A monocular camera-based static semantic map construction method is constructed to convert 3D point cloud maps into 2D maps using projection.
A localization method based on semantic map matching is constructed to align the local map created during localization with the pre-built global semantic map to obtain the current location information.
A method of fusing visual odometry and map matching localization results is constructed to reduce the disadvantages of each of the two localization methods by fusing the two different localization information.

2. METHODS AND EXPERIMENTS

2.1 Semantic Map Construction

Firstly, the semantic segmentation method is used to obtain the semantic information of the image, segment the image and extract the region of interest; then, the geometric constraint information of the feature points in the feature point method SLAM is combined to eliminate the feature points in the dynamic region of the current camera acquisition image to ensure the accuracy of the feature point matching positional estimation; secondly, the meaningless part of the localization process is eliminated, and the semantic information in the region of interest is used to combine with the The semantic map is built by combining the semantic information in the region of interest with the odometer pose results to ensure the validity of the map. Finally, the accumulated errors of key frame poses in the map construction process are corrected by loopback detection or intermittent GNSS information to ensure the positioning accuracy.The semantic map construction process is shown in Figure 2.

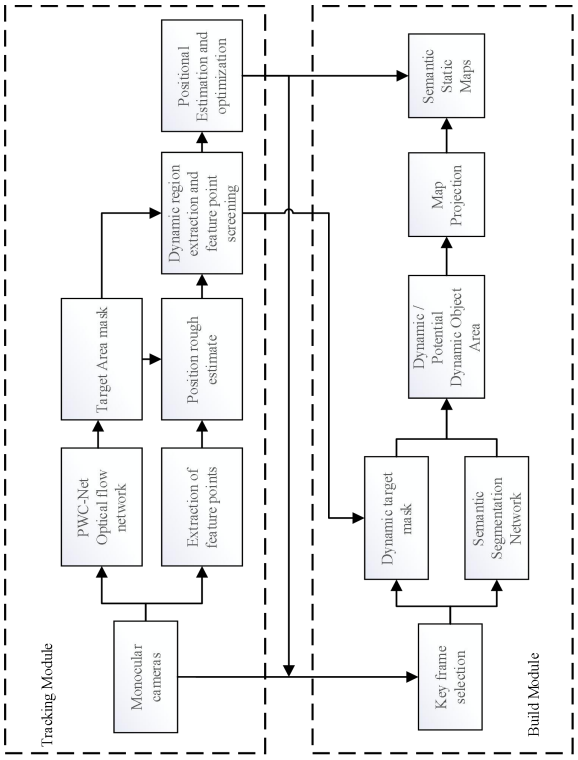

Figure 2. Semantic map construction framework

2.1.1 Semantic Segmentation of Road Environmen： The information contained within the road scene image contains many interfering factors for localization and map building, which need to be filtered by some methods. The semantic information provided by the semantic segmentation based on deep learning can effectively distinguish different semantic regions in the image, and the more common semantic categories include pedestrians, vehicles, sky, buildings, lane markings and so on. Pedestrians and vehicles are common dynamic objects, which have an impact on localization and map building. The feature points detected in the sky can be regarded as invalid feature points, and their depth can be regarded as infinite, which have no practical significance for inter-frame matching and subsequent localization and map building, so the above regions can be segmented out and eliminated. Buildings and road signs and markings can generally be detected as stable feature points, which are necessary features for map building.

The semantic segmentation method based on deep learning first uses the pre-labeled dataset information as the training set, divides the common semantic categories in road scenes and adds semantic labels to the corresponding regions, then uses convolutional neural networks to build a deep learning model, and iteratively trains the training set into the model to obtain a prediction model. Finally, the prediction model is evaluated, and the model is considered to be applicable to semantic segmentation when the input validation set can meet the expected goal when validation is performed. Finally, in practical applications, the image data captured by the input sensor is used to complete the segmentation of the image semantics. At the present stage, deep learning based semantic information acquisition methods have been widely used in SLAM technology (Ganti P and Waslander S L .2018) and (Li P et al. 2018) and (Liang H J et al.2018).

In this paper, we use PP-LiteSeg model for training and prediction of semantic segmentation model. PP-LiteSeg is a lightweight model with a speed of 273.6 FPS on 1080ti model GPU when the average intersection ratio of segmentation accuracy is 72.0, which meets the real-time requirement. The training data uses data images captured through CARLA self-driving car simulator and semantic annotation files, whose semantic labels are divided into 13 categories, unlabeled, buildings, fences, pedestrians, poles, routes, roads, sidewalks, vegetation, cars, walls, traffic signs, and others. Its segmentation effect is shown in the figure 3.

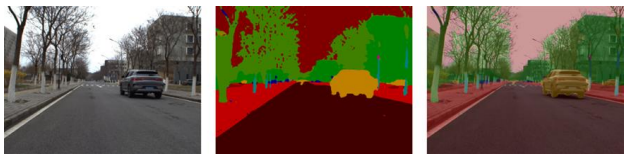

Figure 3. Semantic segmentation of road scenes

2.1.2 Feature Point Processing in Dynamic Environments:

When using SLAM methods for map construction, dynamic objects have a large impact on both the map construction and localization processes. In the feature point visual odometry process, a feature point acquired on the dynamic object, the feature point will be displaced on the pixel coordinate system on the adjacent frame image, if the point is not recognized and rejected by other algorithms, it will make the reprojection error increase in the optimization, and when there are too many feature points in the motion state, it will affect the current odometry calculation results. And in the process of map construction, dynamic objects generally do not want to be represented in the map, such as pedestrians, cars, etc.. Therefore, dynamic objects should be excluded from the map elements. Therefore, the identification and rejection of dynamic areas can improve the positioning accuracy and keep the stability and validity of the map building content.

Acquisition by deep learning semantic segmentation is a very effective means to acquire dynamic regions. For the feature point method SLAM, the feature points on the segmented dynamic objects can be eliminated in the feature point selection stage based on the semantic information provided by the semantic segmentation, and the stable feature points on the non-moving objects on the image can be selected for the localization and map building work. However, distinguishing dynamic feature points only based on the semantic information provided by semantic segmentation will cause some feature points on potential dynamic objects in non-motion state to be eliminated, such as cars in unstarted state, stationary pedestrians, etc. In some scenes, such objects may occupy a larger proportion of the image, resulting in the odometer part of the feature points need to be provided by stationary vehicles, and if they are eliminated instead, it will have an impact on the bit-pose settlement As a result, some other methods are still needed to improve the SLAM method in dynamic environments.

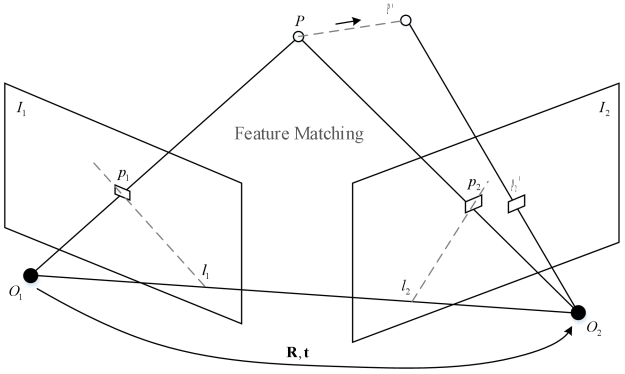

Figure 4. Projection relationship of motion feature points in dynamic environment

To ensure that the camera pose provided in the visual odometry is not affected, geometric constraints are added to the semantic segmentation to determine the dynamic objects that are moving in the current frame. Combining the geometric constraints with the semantic segmentation results can increase the accuracy of the judgment. As introduced in section 2.1.2, there is a certain relationship between the pixel pose and the polar line in two frames, and the object point corresponding to the pixel point in the previous frame has the image point on the polar line in the second frame, which is not satisfied when the object is in motion, as shown in Figure 4. Due to the influence of the camera distortion model, the pixel points of adjacent frames may not perfectly meet the requirements for the polar constraint, so it is necessary to set the threshold value according to the actual situation, and consider the point does not meet the polar constraint when the distance between the pixel point and the polar line is greater than the threshold value. According to the above constraints, firstly, the more stable point is selected according to the semantic information, and the preliminary camera motion pose between two frames is calculated, and the motion feature points are filtered by combining the geometric constraints that should be between this pose and the pixel points. The relationship between the corresponding pixel coordinates of two adjacent frames is:

$$x_2^T F x_1 = \begin{bmatrix} u_2 & v_2 & 1 \end{bmatrix} F \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = 0 \tag{1}$$

Let the vector of polar l_2 lines be $[X,Y,Z]^T$,

$$l_2 = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = F x_1 = F \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} \tag{2}$$

Let the distance between the actual pixel point and the polar line be D. The expression is:

$$D = \frac{\left| x_2^{\mathrm{T}} F x_1 \right|}{\sqrt{\|X\|^2 + \|Y\|^2}} \tag{3}$$

If the distance D is greater than the threshold, the point is considered as not matching the polar line constraint. There are two cases that cause this result; first, these points themselves are

mismatched; second, these points exist on dynamic objects and these points move with the object, causing the mismatch, i.e., such points are dynamic feature points. In both cases, the unsatisfied feature points need to be eliminated. Therefore, all feature points that do not meet the polar constraints can be screened out based on this geometric constraint approach. Combined with the semantic information for judgment, when dynamic feature points are within the detected dynamic region, such feature points are considered as invalid feature points, which can neither be used for visual odometry position estimation nor semantic map construction; when dynamic feature points do not fall within the dynamic region, the object is considered as a non-usually considered dynamic object and is in motion; when static feature points other than those satisfying the geometric constraint are within the dynamic dynamic When the static feature points, except for the geometric constraints, fall within the dynamic region, the dynamic region is considered to contain potentially dynamic objects, and such feature points are only used for pose estimation of visual odometry, but not for semantic map construction.

Figure 5. Feature matching in dynamic environments

Figure 6. Motion object feature point rejection

As shown in Fig. 5 and 6 The feature point detection is performed on two adjacent frames and matched according to the descriptiveness, the hollow circle mark is the detected ORB feature point, and the connecting line indicates that the pair of feature points has been matched, where dynamic feature point removal is not performed in Fig. 6, and Fig. 7 shows the feature matching algorithm of the visual odometer method in dynamic environment constructed in this paper, which can be seen that the method retains the feature point's extracted on the vehicle at rest, removes the feature point matching on the vehicle in motion, and realizes the reduction of the impact of dynamic objects on the visual odometer.

2.1.3 Map Semantic Information Processing : In most scenarios, the vehicle driving process is a ground-constrained moving process, so it can be considered that the positioning result of the vehicle in the vertical direction is known as the elevation of the road surface, so the vehicle positioning information is the two-dimensional information on the road plane. Therefore, based on the map matching localization method, the vehicle only needs to obtain the two-dimensional coordinates of the current location through the a priori map to know its own location, so the map can be downscaled to make the matching process more concise and reduce the matching computation. And in order to avoid significant differences in feature point extraction in different viewpoints in the sparse point cloud map established by feature point method SLAM, a map building method that projects feature points on the road plane is adopted to establish a more stable two-dimensional semantic map.

The map point cloud projection is performed by post-processing. First, the road surface is determined based on the semantic information of each frame with the ground feature points after triangulation and plane fitting in the visual odometer key frame. Then, the feature points in the environment are screened, and only the valid information, such as indication signs, tree trunks, street lights, building facades and other stable areas, are retained and projected onto the road 2D plane; again, the grid is divided to count the semantic labels of the points falling into each grid, and the semantic information of the grid is given according to the maximum number of semantic labels; finally, according to the camera bit pose corresponding to the key frame, the projection image of adjacent frames is completed Finally, the map is built by matching the projection images of adjacent frames according to the corresponding camera pose.

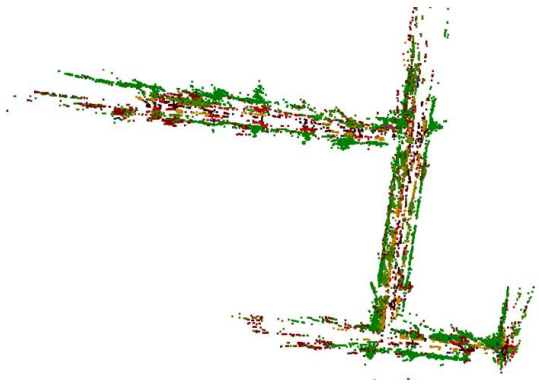

Figure 7. Bird's eye view semantic map

Bird's Eye View (BEV) is a top-down view of the vehicle environment perception map, which can provide the plane location information required for the positioning of the autonomous driving system on the road surface, and the bird's eye view can show the positioning problem more intuitively. The method constructed in this paper projects the point cloud data generated during the map building process combined with its semantic information vertically onto the road plane to generate a semantic map in the form of a bird's-eye view. The projection is shown in Figure 7.

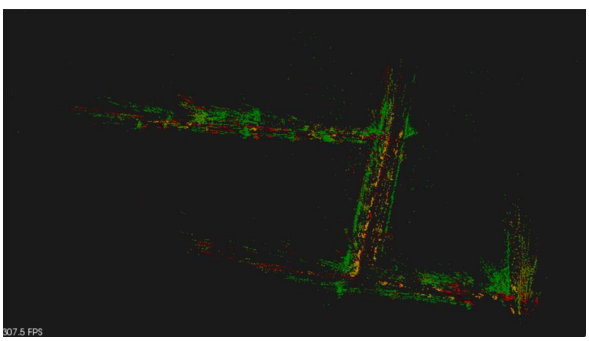

Figure 8. 3D sparse semantic point cloud map

The point cloud map established by the SLAM method of feature points based on vision sensors is often sparse, as shown in Figure 8, and the detection of feature points is affected by the viewpoint, and the feature points detected by the same object under different viewpoints have large differences, resulting in poor results achieved by 3D point cloud matching. Therefore, the feature points in 3D space are projected and converted into 2D form. The road scene has more vertical structures, such as tree trunks, street lights, signage and building facades, etc. The vertical projection of the bird's-eye view enables their positions to be enhanced according to semantic information, as feature points on a pole-like object will be superimposed and projected onto the same area. By the projection method described in the previous section, the feature points in the key frame have been projected into the road plane to obtain a semantic point cloud in the road plane, and now the point cloud obtained is aligned with the world coordinate system by rotation, translation and deflation. The map gains the ability to provide absolute position information.

2.2 Map Matching

After the map format is determined, a map matching-based positioning method needs to be built around the map information to obtain the position information from the map. Compared with other odometry-based positioning methods, the map matching method not only can obtain absolute position information, but also the positioning result has no influence of cumulative error, which is an effective method to solve the positioning in the restricted area of GNSS. Under the premise that the map accuracy is guaranteed, the positioning method needs to meet the positioning accuracy requirements of general vehicle driving. This chapter will elaborate on the constructed semantic map matching-based localization method to improve the accuracy and real-time efficiency of the localization information by fusing the map matching results with the visual odometry poses. Firstly, to solve the scale uncertainty of monocular camera, a method is constructed to recover the scale of the motion trajectory by combining the monocular camera pose provided by the visual odometer with the semantic segmentation information, and using the mounting height of the camera to calculate the depth of the feature points on the road plane, so as to determine the scale of the translation vector between two frames of the camera, and thus recover the scale of the whole motion trajectory; then, according to the map construction Then, the monocular camera data are processed to obtain a local point cloud map with semantic information, and then in the process of matching, the map is matched using the semantic constraint ICP algorithm, using the visual odometry poses as the initial poses for iterative calculation, and using the semantic information constraint in the nearest point selection, and finally, the final positioning results are obtained by fusing the matching positioning results with the visual odometry poses.

2.2.1 Monocular Visual Scale Recovery: In the monocular vision odometry method, the initialization process obtains the bit pose of the adjacent two camera frames by calculating the essential matrix, and the depth information is calculated by triangulation based on the adjacent two frames to recover the spatial position of the feature points, however, since the translation vector solved in the essential matrix is a unit vector. Therefore, only the shape of the object can be determined but not its size attribute can be obtained, as shown in Figure 9. Therefore, the magnitude of the camera translation vector obtained at each initialization of the monocular vision odometer is not the true distance.

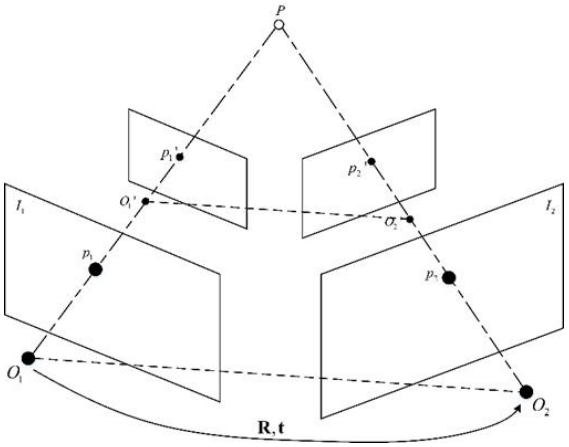

Figure 9.Monocular visual scale uncertainty problem

In order to obtain accurate camera trajectories, it is necessary to determine the depth values of the pixel points in the image corresponding to the depth in 3D space. Therefore, a scale recovery method is constructed in this paper to perform the initial scale calculation by selecting a stably identified and scale-invariant ground as a marker in combination with the camera poses.The semantic information provided in the semantic segmentation can segment the pixel points in the image that represent the road surface, and the relative poses of the camera can be obtained in the essence matrix during the initialization of the visual odometer. At this point, it is only necessary to know the coordinate values of several ground points within the road in the camera coordinate system, and by fitting the road plane through these points, the distance from the camera optical center position to the ground point within the camera coordinate system can be solved using the point-to-plane distance formula. The real coordinate scale can be obtained by solving the distance with the actual camera height value. Thus, the scale of the whole motion process can be recovered.

According to the camera height and the current position of the camera, the coordinates of ground feature points in the world coordinate system can be obtained. This way can compensate for the missing depth during monocular camera measurement, and then recover the scale of the translation vector during the initialization of the odometer, so that the established map scale can be recovered. In order to avoid the influence of measurement errors as much as possible, the ground points should be selected to avoid too far and should be far from the segmentation edge, and the ground feature points should be plane fitted to remove outliers to ensure accuracy. As shown in Figure 10, where the left image is the semantic segmented road plane, the right is the feature point extraction, and the yellow box line is the ground feature points selected to fit the road plane.

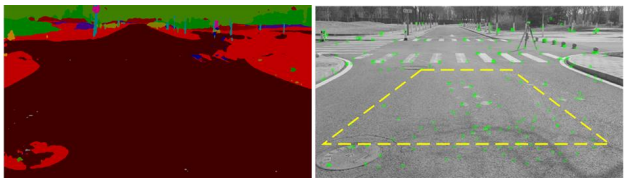

Figure 10. Ground feature point selection

2.2.2 Map Matching Based on Semantic Information： When the positioning end acquires the processed semantic map, it needs to do the same processing of the visual sensor data first, and build a local map by projection, combining multiple frames of images, and match the local map with the previously established global semantic map to obtain the precise position of the local map, and then get the bit position pose when the vehicle acquires the image.

In order to improve the matching effect it is necessary to build a local map to avoid the difficulty of matching ground marker points containing too little information in a single frame and the semantic feature points above the road surface are too sparse leading to the loss of semantic meaning due to the low percentage of their semantic labels in the projection grid, the local map is built at intervals, firstly by moving the distance constraint, when the distance exceeds a threshold, the map built in that time is used to match with the global semantic map, then the matching score is used to control the size of the local map. matching, and then control the size of the local map by the matching score, when the matching score is too low, the result of this matching is not saved and the local map is extended to the next distance threshold moment. However, when the moving distance exceeds the maximum value, the expected matching score is still not reached, and the localization is considered to have failed.

2.2.3 Fusion of Semantic Maps and Visual Odometers for Localization： The previous section describes how to perform semantic map-based matching to obtain the location and pose information of the current collection information in the semantic map to achieve the positioning function, but due to the sparsity of the ground marker feature points and ground markers, for example, in some areas only lane lines exist within the road surface and there are no obvious indication markers, which will make the positioning result in the direction of travel constraint reduction, the accuracy of the map matching positioning result decreases sharply, requiring This results in discontinuous map-based matching results. Therefore, it is necessary to combine the visual odometry to provide continuous inter-frame transformation positional relationships to obtain global and continuous localization results.Due to the effect of systematic error and chance error, it is necessary to fuse the positions acquired by the two position acquisition methods.

Using the nature of the rotation matrix and translation vector, so that the sliding window contains n minus m bit position information, then the i-th($m \le i \le n$) bit position and the current bit position relationship is as follows:

$$\begin{cases} R_n^{\ W} = R_{i,n}^{w} R_i^{w} \\ t_n^{\ w} = R_{i,n}^{w} t_{i,n}^{w} + t_i^{w} \end{cases} \tag{4}$$

where $R_{i,n}^{\ w}, t_{i,n}^{\ w}$ is the rotational translation transformation parameter from the first i-th($m \le i \le n$) positional transformation to the current n-th positional transformation, respectively. Combined with the visual odometry results:

$$\begin{cases} R_{i,n}^{w} = \prod_{s=i}^{n} R_s^{\ c} \\ t_{i.n}^{w} = \sum_{s=i}^{n} R_s^{c} t_s^{c} \end{cases} \tag{5}$$

Let the observed value of the rotational translation parameter of the n-th map matching localization result position in the world coordinate system be $\widehat{R_n^W}, \widehat{t_n^W}$ Then the optimization problem can be constructed:

$$R_{i,n}^{\ W*}, t_{i,n}^{\ W*} = \arg\min \sum_{i=m}^{n-1} \left(\left\| R_{i,n}^{\ W} t_m^{\ C} + t_m^{\ W} - \widehat{t_n^W} \right\|_2 + \left\| (R_{i,n}^{\ W} R_m^C \widehat{R_n^W}^{\mathrm{T}} - I) \right\|_2 \right) \tag{6}$$

The positions obtained by the two position acquisition methods need to be fused because of the systematic and chance errors. In order to calculate the optimal position of the vehicle, a sliding window is designed so that it contains several positions of the map-matched positioning results in the world coordinate system, and all the positioning results in the window are optimized.

3. PRECISION EVALUATION

First, for the relative accuracy verification, this paper verifies that the relative accuracy of the constructed positioning method should be kept within 0.2m (Liu et al.2018) by the data collected by LIDAR. The LIDAR sensor performs ranging by actively emitting a signal source, and its accuracy is relatively high, usually its range accuracy is centimeter per hundred meters. The specific steps are to first complete the semantic map construction using the map construction method, and determine the road edge location based on the semantic points. Then, the localization is performed by the map matching method constructed in this chapter, and the position of the localization trajectory in the map is determined. Secondly, the single-frame LiDAR road edge line position is extracted using the LiDAR collected simultaneously in the localization process, and the LiDAR data is aligned with the localization key frames according to the time stamps, and finally, the relative positions of the localization results in the road are compared by the aligned coordinate system. Within this framework, this paper compares and analyzes the effects of different semantic map division accuracy on localization, and obtains the best map processing parameters by comparing the effects of different semantic grid division sizes on relative accuracy during the semantic processing stage.

Second, for the absolute accuracy verification, this paper uses the image data of the public dataset for verification, so that the absolute positioning accuracy is held within 1m (Fischler M A and Bolles R C.1981). Firstly, in the dataset, repeatedly passing road sections are selected as the experimental data, and the positioning results are calculated using the first passing images as the map building data and the subsequent passing images as the positioning data. And according to the true value of the trajectory provided in the dataset and the localization results in the method of this paper are compared and analyzed, while adding only the visual odometry method in ORB_SLAM2 is used for comparison to verify the effectiveness of the method of this paper.

Some line routes in the Kitti dataset 00 sequence were selected for the experiment. In the 00 sequence, 388 frames to 938 frames of images are used as map construction data, and 3398 frames to 3842 frames of images are used as localization data. The complete data of the sequence all constitute the loopback, and the bit pose correction of the map construction using data of key frames is performed by loopback detection.

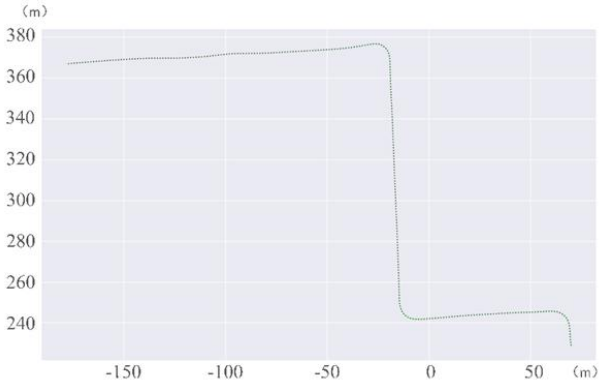

Figure 11. Kitti dataset 00 sequence trajectory

According to the trajectory true value data provided by the dataset, the length of the line in the 00 sequence part of the dataset is 385.328 m and the length of the line in the 00 sequence part is 284.62 m. According to the experimental results in the previous paper, 0.1m resolution maps were used in the lines. The positioning results were analyzed and visualized using the evo tool, and the absolute error values were calculated by comparing the ground truth with the positioning results At the same time, the visual odometer method without fused map matching results was used as a comparison term to calculate the absolute error of the positioning results. In this experiment, the visual odometer trajectory has been aligned with the true value of the trajectory by rotation and translation operations to reduce the influence of the initial angle and position on the accuracy. The absolute positioning error results of the 00 sequence are shown in Figure 12. After that, the absolute error value of the line is then counted, and the absolute error is divided by interval according to the error size, and the distribution is counted to indicate the stability of the positioning results, where the error distribution of 00 sequence is shown in Figure13.

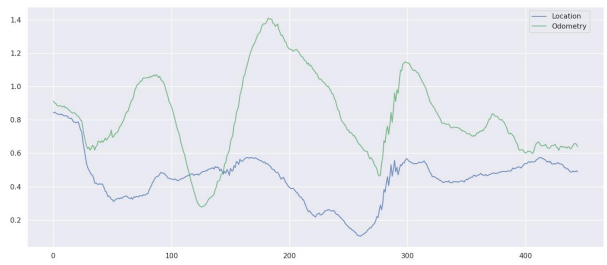

Figure 12. Kitti dataset 00 sequence absolute position error

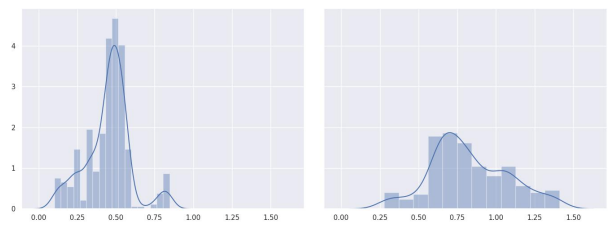

Figure 13. Distribution of Kitti dataset 00 sequence absolute position error

As shown in Figure 13, the ground truth of the trajectory is used as the comparison, and the length of the trajectory route is used as the x-axis, and the positioning error value of the results of this paper and the visual odometry results is used as the y-axis, it can be seen that the absolute positioning accuracy of the map matching-based positioning method is better than that of the visual odometry results in this data set. The error distribution of the map matching method is smaller and more concentrated than that of the visual odometer, with a median error of 0.47 m for the map matching method and 0.78 m for the visual odometer, so the map matching method effectively corrects the offset of the trajectory route in the visual odometer and reduces the cumulative error of positioning during driving.

4. CONCLUSIONS

In this paper, accuracy verification experiments are designed for the reliance on absolute and relative positions during vehicle driving. First, the relative positioning accuracy of the positioning method is verified by comparing the positioning results in the lateral direction in the road through the camera and LiDAR equipped on the data collection platform built by the team. Second, the absolute localization accuracy of the semantic map matching-based localization algorithm is evaluated by comparing the localization results in the public dataset with their ground truth values, and adding visual odometry results for comparison to prove the effectiveness of the localization method in this paper. The positioning method constructed in this paper is comprehensively evaluated by two experiments.

5. ACKNOWLEDGEMENTS

This research was funded by the National Natural Science Foundation of China (Grant Numbers 42201483) and the China Postdoctoral Science Foundation (Grant Numbers 2022M710332) and Research on Orthophoto Generation.

6. REFERENCES

Campos C , Elvira R , JJG Rodríguez, et al. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM[J]. 2020.

Dong G , Yan Y , Shen C , et al. Real-Time High-Performance Semantic Image Segmentation of Urban Street Scenes[J]. IEEE, 2021(6).

Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 24(6):381-395.

Fleet D , Pajdla T , Schiele B , et al. [Lecture Notes in Computer Science] Computer Vision – ECCV 2014 Volume 8690 || LSD-SLAM: Large-Scale Direct Monocular SLAM[C]// 2014:834-849.

Ganti P , Waslander S L . Visual SLAM with Network Uncertainty Informed Feature Selection[J]. 2018.

Hong Y , Pan H , Sun W , et al. Deep Dual-resolution Networks for Real-time and Accurate Semantic Segmentation of Road Scenes[J]. 2021

Jeong J , Cho Y , Kim A . Road-SLAM : Road marking based SLAM with lane-level accuracy[C]// IEEE. IEEE, 2017:1736-1473.

Li P , Qin T , Shen S . Stereo Vision-based Semantic 3D Object and Ego-motion Tracking for Autonomous Driving[J]. Springer, Cham, 2018.

Liang H J , Sanket N J , C Fermüller, et al. SalientDSO: Bringing Attention to Direct Sparse Odometry[J]. IEEE Transactions on Automation Science & Engineering, 2018.

Liu Jingnan, Wu Hangbin, Guo Chi, Zhang Hongmin, Zuo Wenwei, Yang Cheng. The progress and reflection of high precision road navigation map[J]. China Engineering Science,2018,20(02):99-105.

Peng J , Liu Y , Tang S , et al. PP-LiteSeg: A Superior Real-Time Semantic Segmentation Model[J]. 2022.

Qin T , Zheng Y , Chen T , et al. RoadMap: A Light-Weight Semantic Map for Visual Localization towards Autonomous Driving[J]. 2021.

Tong, Qin, Peiliang, et al. VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator[J]. IEEE Transactions on Robotics, 2018.