Multi-Sensor-Fusion Localization Assisted By Pole-Like Objects In Long Range GNSS Degraded Environment

Yanhai Ma, Shuai Wang, Xudong Wang, Yongliang Wang
Riemann Lab, 2012 Laboratories, Huawei Technologies Company, Ltd., Beijing 100095，China

Biography

Yanhai Ma received the doctor's degree from Beihang University, China in 2015. He is currently working as a principle engineer with Riemann Lab, 2012 Laboratories, Huawei Technologies Company Ltd. His main research interests include multi sensor fusion state estimation, perception in autonomous driving, crowd-sourced high-definition map.

Shuai Wang received the master's degree in Information Technology from University of Melbourne, Melbourne, Australia, in 2018. He is currently working as a senior engineer with Riemann Lab, 2012 Laboratories, Huawei Technologies Company Ltd. His research interests include deep learning, visual recognition in autonomous driving.

Xudong Wang received the master's degree in computer science and technology from Xi'an Jiaotong University, Xi'an, China, in 2021. He is currently working as a senior engineer with Riemann Lab, 2012 Laboratories, Huawei Technologies Company Ltd. His main research interests include deep learning, 3D object detection and vision-centric perception.

Yongliang Wang received the master's degree from the School of Electronic and Information Engineering, Xi'an Jiaotong University, China, in 2010. He is currently worked as a principle engineer with Riemann Lab, 2012 Laboratories, Huawei Technologies Company Ltd. His research interests include high-resolution localization and high-resolution map.

Abstract

The global position precision of High-definition (HD) Map mainly relies on GNSS Real-Time Kinematic (RTK) positioning. RTK solution cannot be fixed in long urban canyon or long tree-lined roads for the non-line-of sight (NLOS) of signal and multi-path effect of signal reflection, consequently, the localization error in relative localization system such as IMU/wheel encoder, visual-inertial odometry (VIO), Lidar-inertial odometry (LIO) will drift dramatically. To improve the large drift error under long range GNSS denied environment, a multi-sensor-fusion localization algorithm assisted by pole-like objects registration is proposed in this paper. Firstly, an efficient relative pose observation model for two distant states are built by registration of pole-like object is proposed, in which an effective registration algorithm between pole-like objects detected by point-cloud and that detected by semantic segmentation on image plane algorithm is proposed. Secondly, a multi-sensor-fusion localization system which incorporate pole-like objects registration is proposed to get more precise global positioning result. The evaluation of the method proposed in this paper was carried out by the data acquired from a HD Map acquisition vehicle, the result shows that relative positioning precision improved from 0.5% distance to 0.18% distance error, the max absolute positioning error improved from 0.42 meter to 0.18 meter whenever the RTK is unavailable in a tracking distance of 500 meters.

I. INTRODUCTION

HD Map is of great importance in autonomous driving which is usually used for autonomous driving car's localizing and path planning. The global position accuracy of ten centimeters' level is required when constructing HD Map, which is usually obtained by GNSS-RTK or offline GNSS carrier phase differential positioning [1]. However, there almost exist many hundreds of meters' urban canyon and tree lined roads, the GNSS RTK cannot be fixed and the localization error is of meter's level or ten meter's level, which is induced by NLOS of satellite signal and multi-path effect [2,3]. To avoid this problem, one of the practical methods is to change HD Map construction scheme by using data mining from aerial image, which avoid the localization error of GNSS receiver mounted on vehicle [4]. However, the high definition map elements such as traffic signs, traffic lights and slope cannot be extracted

 https://doi.org/10.33012/2022.18548

from aerial image for the shooting angle of camera is from top to down. To overcome this problem, there are always a procedure that mapping the map from aerial data source to that from land vehicle data source, which is laborious and costly. Other methods to solve this problem include improving the precision of GNSS solution itself, multi-sensor-fusion localization and correlation localization with environments' features.

As for GNSS multi-path effect mitigation, on the one hand, receiver signal processing technology is usually utilized, on the other hand, by designing GNSS antenna as right-hand circularly polarized (RHCP) can restrain the reflected signal [5, 6]. As for NLOS error, the detection can be enhanced by using an antenna array to measure angle of arrival or using a dual-polarization antenna can enhance the NLOS detection, however, additional hardware is not always practical [7]. Another means to improve the GNSS precision is using 3D-mapping aided (3DMA) technology [8, 9]. Ranging-based GNSS positioning aided by 3d-mapping and shadow matching can improve positioning accuracy along street and across street separately [10, 11], they are integrated by a different weighting, however, the error is still of meter's level, which cannot meet the HD Map's requirement.

Multi-sensor-fusion localization scheme often use some kind of combination of GNSS, IMU, wheel encoder, camera and Lidar. GNSS provide the global position and other sensors provide the relative localization [12, 13]. The data from different sensors can be fused by Extended Kalman Filter (EKF) or non-linear optimization. The sliding-window based optimization scheme achieved better accuracy than filter-based algorithm in multi sensor fusion processing for that it constrains more relevant states to accord with multi sensors' observations in a continuous time period. The propagation drift of inertial navigation is low in short time span. However, it will drift fast in long integration time. Wheel encoder can provide a stable measurement of instantaneous velocity. So that the distance of travelling is with low error, the drawback of IMU/wheel encoder fusion is that the attitude drift cannot be restrained. Visual-Inertial Odometry (VIO) is a fashionable fusion scheme nowadays because that both the position error and attitude drift are low, for that the feature points near camera can provide position constraints with good quality while the point extracted far from camera provide attitude constraints with good quality [14, 15]. However, the feature points far away from camera is rare and reconstruction is not very accurate, which reduces the odometry's precision, in addition, with the expansion of sliding window, the computation burden is heavy. So that the sliding-window is often narrow and it only include a few states which are consecutive is time span and neighbouring in space. The old state will slide out of the window with time transition, so that the direct correlation relationship between the state with large drift which transit a long time and previous state with small drift cannot be obtained, so that the large drift error cannot be reduced. The similar problem also exists in the lidar odometry, with time transition, the drift error is large, while the same objects cannot be observed by a state in history, for that the lidar scan is sparse, the object that is far from the state will not be observed, so that the direct observation cannot be built in Lidar odometry [16-18].

Environmental correlation methods usually register the objects sensed by on-board sensors such as camera and Lidar with that of pre-built map to avoid drift under GNSS denied situation. Inspired by this method, if an object can be observed by both the sensor whose state is with large drift produced by relative localization proceeding with time and the sensor whose state is with small drift when the relative localization algorithm starts, the large drift error can be rectified by registration this object in two sensor's output. To meet this requirement, the object need to exist long term and easy be distinguished by a sensor far from it, at the same time, the object can be extracted precisely when observed by sensor nearby.

In this paper, to overcome the localization drift under long GNSS denied environment, multi-sensor-fusion localization assisted by pole-like objects is proposed. There are two main contributions. Firstly, an efficient relative pose observation model for two distant states are built by registration of pole-like object is proposed. Secondly, IMU/Wheel Encoder/GNSS fusion localization assisted by pole-like objects is proposed.

The rest of this paper are organized as follows: Section 2 introduced the pole-like objects semantic segmentation based on both point cloud and image. Section 3 describes the construction of relative pose observation model for two distant states by pole-like registration. Section 4 describes the IMU/Wheel encoder/GNSS fusion localization algorithm, Section 5 describes the field tests and experimental results. Section 6 offers conclusion.

II. POLE-LIKE OBJECTS SEMANTIC SEGMENTATION

HD Map acquisition vehicle are usually equipped with cameras, LIDARs, and other sensors. In this paper, the pole-like objects need to be extracted precisely from both point clouds and images. Semantic segmentation is the best method to extract pole-like objects from images and point clouds. Although these semantic segmentation methods are relatively mature, we still give a basic introduction to the methods used here.

The technology of objects' semantic segmentation from images we used is High-Resolution Network (HRNet) [19]. The network is lightweight and efficient. High-resolution features with rich semantics and precise spatial position can be learned by HRNet. This benefit from the network structure design that higher resolution features extracted from those shallow layers are kept and fused together.

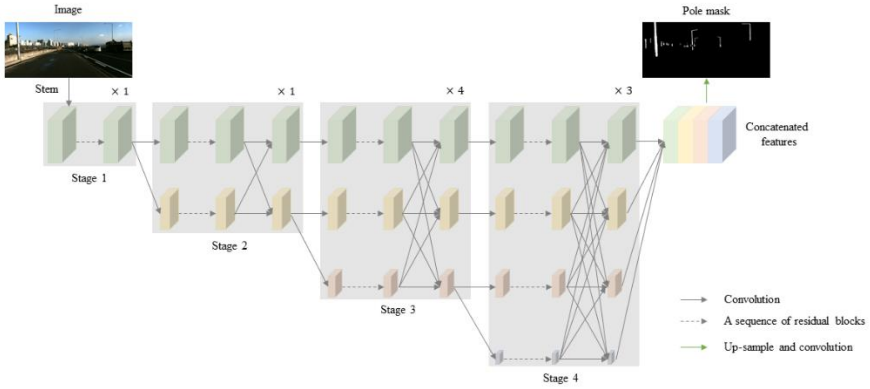

Figure 1 *An overview of the HRNet semantic segmentation network.*

In above figure, the four stages of HRNet generate stacks of multi-resolution feature maps. These feature maps are concatenated and mixed, then up-sampling to image resolution. A simple convolution is finally used to produce pole mask [20].

Pole-like objects segmentation from point cloud is hard in single laser scan, because the laser beams spread with an angle which makes the point cloud of objects far from sensor sparse. Sometimes, LIDAR in HD Map acquisition vehicle is single ring or double ring, which makes the segmentation impossible in single laser scan. To overcome this drawback, point clouds of multi scans are transformed to obtain a local map by poses produced by IMU/Wheel encoder fusion in a short time, for that its drift in short time is tiny. Subsequently, the pole-like objects' segmentation are executed by a three-stage method. Firstly, a set of linear voxels including pole-like object candidates are generated by the voxel-based shape recognition [21]. Specifically, point clouds space are divided into regular 3d voxels. Then, by using principal component analysis (PCA) to analyze the dimension of voxel, the point cloud shape can be divided into three types: linear, planar and spherical. Secondly, those linear voxels are clustered to form spatially independent individuals. According to the characteristics of height and thickness of poles, some individuals which are obviously not pole-like will be filtered out. At present, remainder individuals may include lamp pole, signboard pole, electric pole, tree and so on. Tree pole is not the pole-like we want, so we need to classify tree and artificial rod in the second stage. It is noticed that the point cloud on the top of the tree are scattered, while the point cloud on the artificial rod was concentrated. Based on this clue, we can accurately classify these two types of pole-like objects [22]. Finally, min-cut based segmentation are used to divide the point cloud into foreground and background. The foreground is the artificial pole-like that we want [23].

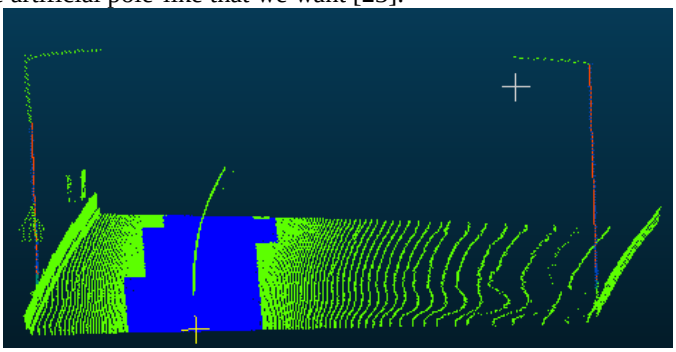

Figure 2 *Pole-like segmentation from point cloud*

In figure 2, the segmentation of pole-like objects such as lamp pole and signboard pole from point cloud are shown in red color, other color is background.

III. RELATIVE POSE OBSERVATION MODEL FOR TWO DISTANT STATES

The localization error of multi-sensor-fusion localization system increases with time under no valid position observation from GNSS, to alleviate this problem, relative pose observation for two distant states can be built so that drift error between them will be reduced. For that the pole-like objects such as lamp pole and signboard pole exists ubiquitously in city urban area and changed rarely. Furthermore, range measurement of laser is precise, the registration of pole-like objects is processed as follows: i) IMU/wheel

encoder fusion will be executed firstly to obtain high-frequency relative pose output, then, both camera and Lidar scan's pose are interpolated from IMU/wheel encoder fusion output. ii) Tracking trail is divided into many segments by a long travelling distance. The start pose of each segment is synchronized with image and is expressed as T_{is}, correspondingly, the end pose of each segment is synchronized with Lidar scan and is expressed as T_{ie}. iii) Local maps associated to each T_{ie} are established by point cloud transformation of those scans collected after T_{ie}, subsequently, the pole-like objects semantic segmentation result can be obtained through the 3-stage method introduced in II iv) Pole-like objects are projected to image at T_{is} and the projection error is established. To accelerate the optimization process, the projection error is established in advance. Once the semantic segmentation of pole are obtained, the error map will be established by building chamfer distance error for each pixel, the error in foreground pixel is zero, the error in background pixel is calculated to nearest foreground pixel. The whole process of projecting a point in local map associated to T_{ie} into image associated to T_{is} is as follows:

First, the feature point in local map is transformed to world frame, see equation (1).

$$^{\mathbf{W}}\mathbf{P_f} = \mathbf{T_L^W}\ ^{\mathbf{L}}\mathbf{P_f} \tag{1}$$

Where $^{\mathbf{L}}\mathbf{P_f}$ is a feature point expressed in local map frame, $\boldsymbol{T_L^W}$ is the transformation of world frame with respect to local point cloud map at each segment end time. $^{\mathbf{W}}\mathbf{P_f}$ is the feature point expressed in world frame, as we proposed in former, the local point cloud map is accumulated from the time of T_{ie}.

Second, the feature point in local point cloud map is projected to the camera frame captured at T_{is},

$$^{\mathbf{C}}\mathbf{P_f} = \mathbf{T_I^C T_W^I T_L^W}\ ^{\mathbf{L}}\mathbf{P_f} \tag{2}$$

Where $\mathbf{T_W^I}$ is the transformation of IMU frame with respect to world frame at each segment start time, and $\mathbf{T_I^C}$ is the transformation of camera frame with respect to IMU frame, which is constructed from extrinsic parameters between this two frames. $\mathbf{T_L^W}$ is with large drift, the measurement of drift should be measured by registration of poles between that in local map frame adhered to T_{is} and that in image frame adhered to T_{ie}.

To explicitly express the orientation and translation, equation (2) is expanded as follows:

$$^{\mathbf{C}}\mathbf{P_f} = \mathbf{R_I^C R_w^I R_L^W}\ ^{\mathbf{L}}\mathbf{P_f} + \mathbf{R_I^C R_w^I}\ ^{\mathbf{W}}\mathbf{P_L} + \ ^{\mathbf{C}}\mathbf{P_w} \tag{3}$$

Where $\mathbf{R_L^W}$, $^{\mathbf{W}}\mathbf{P_L}$ is the rotation matrix and translation of world frame with respect to local point cloud map frame at each segment end time. $\mathbf{R_W^I}$ is the rotation matrix from world frame to IMU frame at each segment start time. $\mathbf{R_I^C}$ is the rotation matrix from IMU frame to camera frame, and which is sourced from extrinsic parameters. $^{\mathbf{C}}\mathbf{P_w}$ is the translation of world frame center with respect to camera frame at each segment start time.

Jacobian matrix of $^{\mathbf{C}}\mathbf{P_f}$ with respect to $\mathbf{R_L^W}$ and $^{\mathbf{W}}\mathbf{P_L}$ can be derived from equation (3) as:

$$\begin{bmatrix} \frac{\partial\ ^{\mathbf{C}}\mathbf{P_f}}{\partial \boldsymbol{\phi}_{\mathbf{L}}^{\mathbf{W}}} & \frac{\partial\ ^{\mathbf{C}}\mathbf{P_f}}{\partial\ ^{\mathbf{W}}\mathbf{P_L}} \end{bmatrix} = \begin{bmatrix} -\mathbf{R_I^C R_w^I R_L^W}(\ ^{\mathbf{L}}\mathbf{P_f})^{\wedge} & \mathbf{R_I^C R_w^I} \end{bmatrix} \tag{4}$$

Where $\boldsymbol{\phi}_{\mathbf{L}}^{\mathbf{W}}$ is rotation vector represents the perturbation of rotation matrix $\mathbf{R_L^W}$, $(\ ^{\mathbf{L}}\mathbf{P_f})^{\wedge}$ is the skew-symmetric matrix of $^{\mathbf{L}}\mathbf{P_f}$.

Then, the point in camera frame are projected to the normal sensing plane (NSP), this process can be expressed as:

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}_{\boldsymbol{nsp}} = \frac{1}{z_c} \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{x_c} \\ \boldsymbol{y_c} \\ \boldsymbol{z_c} \end{bmatrix} \tag{5}$$

Where x_c, y_c, z_c represents the point's coordinates in camera frame. Correspondingly, jacobian of the point in NSP with respect to that in camera frame can be obtained as:

$$\frac{\partial [x_{nsp} \quad y_{nsp}]^T}{\partial [x_c \quad y_c \quad z_c]^T} = \begin{bmatrix} \boldsymbol{1/z_c} & \mathbf{0} & \boldsymbol{-x_c/z_c^2} \\ \mathbf{0} & \boldsymbol{1/z_c} & \boldsymbol{-y_c/z_c^2} \end{bmatrix} \tag{6}$$

For that the distortion exists in imaging process, the point in NSP need to be transformed to the distorted normal sensing plane (DNSP).

$$\boldsymbol{r^2 = x^2 + y^2} \tag{7}$$

Where x, y represents the coordinates of point in NSP, r represents the radial distance from center of NSP.

$$\boldsymbol{x_{dnsp} = (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) x + 2 p_1 xy + p_2 (r^2 + 2x^2)} \tag{8}$$

$$\boldsymbol{y_{dnsp} = (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) y + 2 p_2 xy + p_1 (r^2 + 2y^2)} \tag{9}$$

Correspondingly, the jacobian of point in DNSP with respect to that in NSP is obtained as follows:

$$\frac{\partial x_{dnsp}}{\partial x_{nsp}} = \boldsymbol{1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + 4x^2 r(k_1 + 2k_2 r^2 + 3k_3 r^4) + 2p_1 y + 4p_2 x + 4p_2 rx} \tag{10}$$

$$\frac{\partial x_{dnsp}}{\partial y_{nsp}} = \boldsymbol{4xyr(k_1 + 2k_2 r^2 + 3k_3 r^4) + 2p_1 x + 4p_2 ry} \tag{11}$$

$$\frac{\partial y_{dnsp}}{\partial x_{nsp}} = \boldsymbol{4xyr(k_1 + 2k_2 r^2 + 3k_3 r^4) + 2p_2 y + 4p_2 rx} \tag{12}$$

$$\frac{\partial y_{dnsp}}{\partial y_{nsp}} = \mathbf{1} + \boldsymbol{k_1 r^2} + \boldsymbol{k_2 r^4} + \boldsymbol{k_3 r^6} + \mathbf{4}\boldsymbol{y^2 r}(\boldsymbol{k_1} + \mathbf{2}\boldsymbol{k_2 r^2} + \mathbf{3}\boldsymbol{k_3 r^4}) + \mathbf{2}\boldsymbol{p_2 x} + \mathbf{4}\boldsymbol{p_1 y} + \mathbf{4}\boldsymbol{p_1 r y} \tag{13}$$

Finally, the point in DNSP is project to the image plane,

$$\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} = \begin{bmatrix} \boldsymbol{f_x} & \boldsymbol{s} & \boldsymbol{c_x} \\ \mathbf{0} & \boldsymbol{f_y} & \boldsymbol{c_y} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}_{\boldsymbol{dnsp}} \tag{14}$$

where $\boldsymbol{f_x}, \boldsymbol{f_y}$ represent the focal lengths, $\boldsymbol{c_x}, \boldsymbol{c_y}$ represent the camera principal point, and s represent skew coefficient. Jacobian of point in image plane with respect to that in DNSP is obtained as follows:

$$\frac{\partial [u \quad v]^T}{\partial [x_{dnsp} \quad y_{dnsp}]^T} = \begin{bmatrix} \boldsymbol{f_x} & \boldsymbol{s} \\ \mathbf{0} & \boldsymbol{f_y} \end{bmatrix} \tag{15}$$

The point in local point cloud map at end of each segment can be projected to the image captured at start of each segment.

Figure 3 *Projection of pole-like objects*

In figure 3, the left is the overview of pole-like objects' projection, the right is the local enlarged view, where the green points represent projection of the pole-like objects from local point cloud map to image. Obviously, due to the fusion drift of IMU / wheel encoder, the projection of pole-like objects in the local map do not coincide with that in the image. To adjust this drift, the projection error should be measured. Because there is no obvious texture on the pole-like objects to extract distinguishable feature point, direct projection error of point on pole-like objects cannot be obtained. To solve this problem, the overall error of whole pole-like object can be modeled by the chamfer distance error, the chamfer distance error of pole-like objects in image is zero. Whenever the point projection location is different from pole-like objects segmentation, the chamfer distance error will be exist. The chamfer distance error map of an image can be shown as follows:

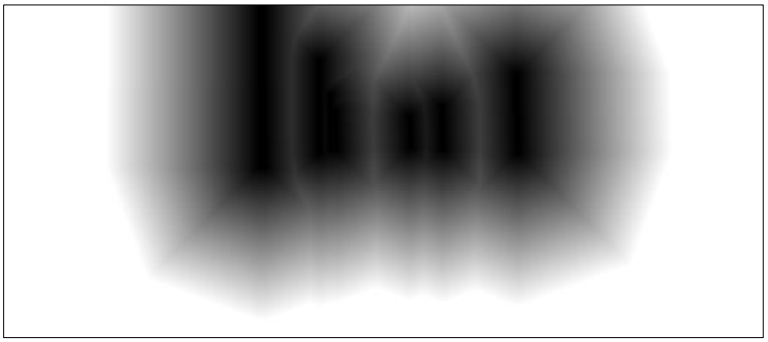

Figure 4 *Chamfer distance error map of pole-like objects segmentation in image*

In figure 4, the dark part represent the chamfer distance error is close to zero, while the bright while part implies the chamfer distance error is large for that it is far from pole-like objects. For that the projection of a point in image is of sub-pixel, the chamfer distance error is obtained by bilinear interpolation of its adjacent four pixels, and jacobian of chamfer distance error with respect to image coordinates can be obtained by bilinear interpolation accordingly.

Relative pose observation for start pose and end pose of a trajectory segment can be obtained by pole-like objects registration adhered to these two poses. The cost function of this problem is composed of all pole-like objects's residual, and which is composed of residual of all sample point.

$$\min_{\boldsymbol{x}} \frac{\mathbf{1}}{\mathbf{2}} \sum_i \left(\sum_j \left\| \boldsymbol{f_j}(\mathbf{x}) \right\|^{\mathbf{2}} \right) \tag{16}$$

Where subindex i and j indicates the index of pole-like objects and sample point on pole-like object. And x represent the state of T_{ie}, by nonlinear optimization, the state of T_{ie} can be adjusted to meet the observation of pole-like objects in image. Correspondingly, the relative pose error between T_{is} and T_{ie} can be calculated.

Figure 5 *Pole-like registration between start state and end state of a trajectory segment*

In figure 5, the right part is the local enlarged view of pole-like registration. It can be found that the pole-like objects in local map are registered to the correct position, accordingly the error of state T_{ie} is reduced.

IV. IMU/WHEEL ENCODER/GNSS/POLE-LIKE OBJECTS FUSION LOCALIZATION

From part 3, the large relative drift error can be measured by pole-like registration, this will reduce the drift error under long GNSS degraded environment. To simply the localization procedure, loosely-coupled fusion scheme is adopted. The state is initialized from IMU/wheel encoder fusion and are constrained by GNSS observation, pole-like registration observation. The system structure of multi-sensor-fusion localization is shown in figure 6.

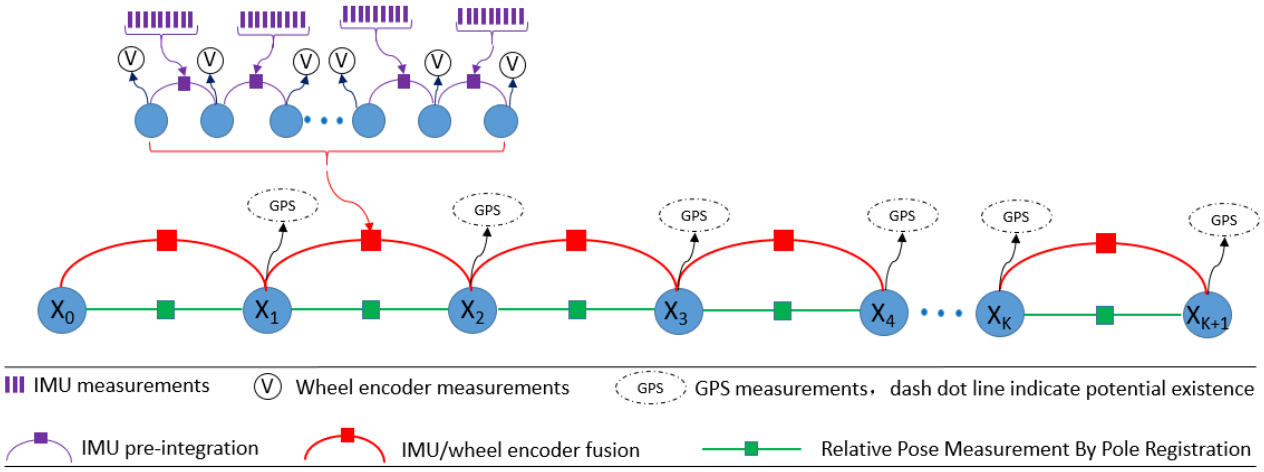

Figure 6 *The system structure of IMU/wheel encoder/GNSS fusion assisted by pole registration*

In figure 6, the system receives input from lidar, camera, IMU, wheel encoder and GPS. Five types of factors are introduced to construct the factor graph: (a) IMU pre-integration factor, which is used in IMU/wheel encoder fusion and system initialization. (b) wheel encoder velocity factor. (c) GPS factor. (d) IMU/wheel encoder fusion factor. (e) Relative pose estimation factor by pole registration.

The IMU pre-integration is a relatively mature technology, detail of its factor are not listed here. Wheel encoder measurement is a scalar, so that the wheel speed is decoupled into vehicle body coordinates by attitude originated from IMU/wheel encoder fusion. Inertial sensor's bias are estimated by wheel speed and GPS measurements. Which is the position observation in WGS-84 coordinates. IMU/wheel encoder fusion factor is the relative pose of two states. Similarly, the relative pose estimation factor by pole registration is of same characteristics.

In practical execution, when registration of pole-like objects between trajectory start and trajectory end, due to the projection of samples on poles in trajectory end may be located out of the image range collected at trajectory start, or due to occlusion by moving

vehicles, last trajectory end are not always taken as the next trajectory start, the next trajectory start is often near last trajectory end in a small range, the gap between current trajectory end and last trajectory start is filled by IMU/wheel encoder fusion results.

V. EXPERIMENTAL EVALUATION

In this paper, we use the data acquired from HD Map acquisition vehicle to verify the proposed algorithm. SZT-R1000 is a vehicle mounted integrated mobile measurement system which is developed by South Surveying and Mapping Group, China. The equipment is composed of high precision three-dimensional laser scanner, GNSS satellite positioning system, inertial navigation system (INS), panoramic camera and control module, highly integrated control module, time synchronization module, etc.

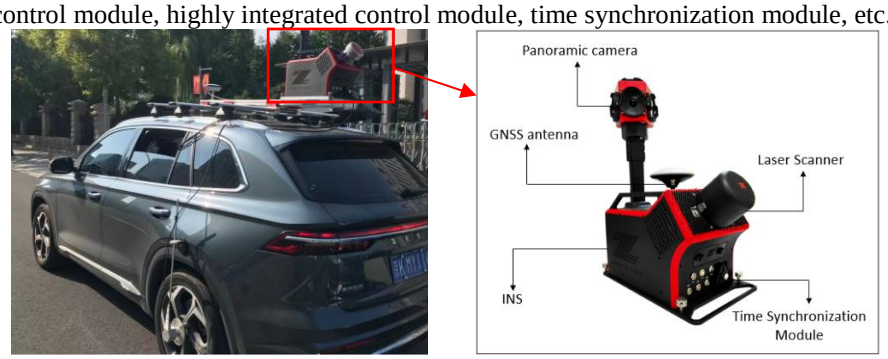

Figure 7 *HDMap acquisition vehicle and data collection system SZT-1000.*

In SZT-1000 data acquisition system, panoramic camera is composed of six HD camera of five megapixel, among them, five cameras are mounted horizontally, and one camera is mounted upward to capture scene in the sky, and all cameras are global shutter, and its sampling frequency is 2 Hz. Laser scanner is single-ring and mounted with angle of 45 degree with respect to vehicle, this setup makes laser ray reaches up to higher buildings, angle resolution of laser scanner is up to 0.001 degree, laser emission frequency is up to 1800 KHz. The IMU in INS is IMU-KVH1750, the bias stability of gyroscope is 0.05 degree/hour, the bias stability of accelerometer is 7.5 mg, its sampling frequency is 200 Hz. GNSS use NovAtel OEM7, which receives multi signal from GPS, GLONASS, Beidou, its sampling frequency is 1 Hz, RTK solutions were obtained from Qianxun FindCM service. The integration of GNSS RTK/INS was used to provide ground truth of positioning.

The test was executed in Xian city, Shanxi Province, China. Test tracking route is shown in figure 8, vehicle stayed stationary at start of track which was marked with red square and ends at position which was marked with yellow triangle in this figure. Total distance of this route is about 1.2Km and the duration is about 160 seconds. To provide the ground truth, there were no tall buildings to block satellite signal, to simulate the environment of city canyon and tree-lined roads, only a few RTK solution were kept to simulate long-term GNSS denied situation.

Figure 8 *Test tracking route in Xian city.*

Firstly, relative pose estimations originated from IMU/Wheel encoder fusion and from IMU/Wheel encoder fusion assisted by pole registration were obtained and compared when there is no precise GNSS-RTK solution. In figure 9, the track originated from IMU/Wheel encoder fusion was drawn in blue and the track from IMU/wheel encoder fusion assisted by pole was drawn in red, while the ground truth was drawn in green. Two area are enlarged to see the comparison explicitly, it was found that the drift of odometry fusion assisted by pole registration is small than that without pole registration assisted.

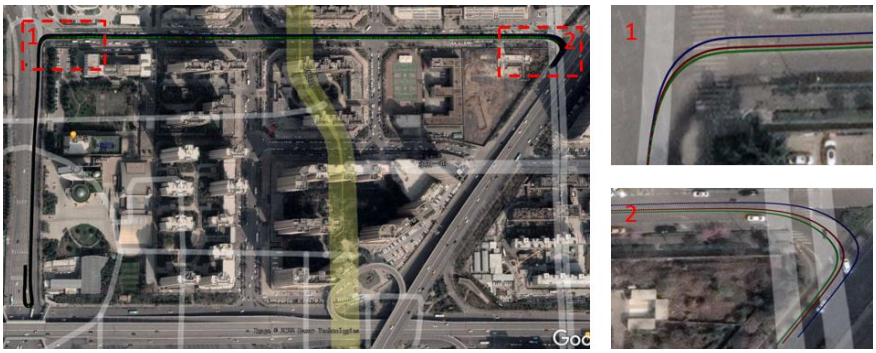

Figure 9 *Relative pose estimation comparison between IMU/Wheel encoder fusion assisted by pole registration and no assistance.*

Figure 10 gives comparison curve of translation error. The localization translation error was showed in a frame where x axis point to right of vehicle, and y axis point to forward of vehicle on initial moving time. It is shown that the translation error increasing direction is consistent with tracking direction of vehicle, which demonstrate that the error mainly came from the scale error of wheel encoder.

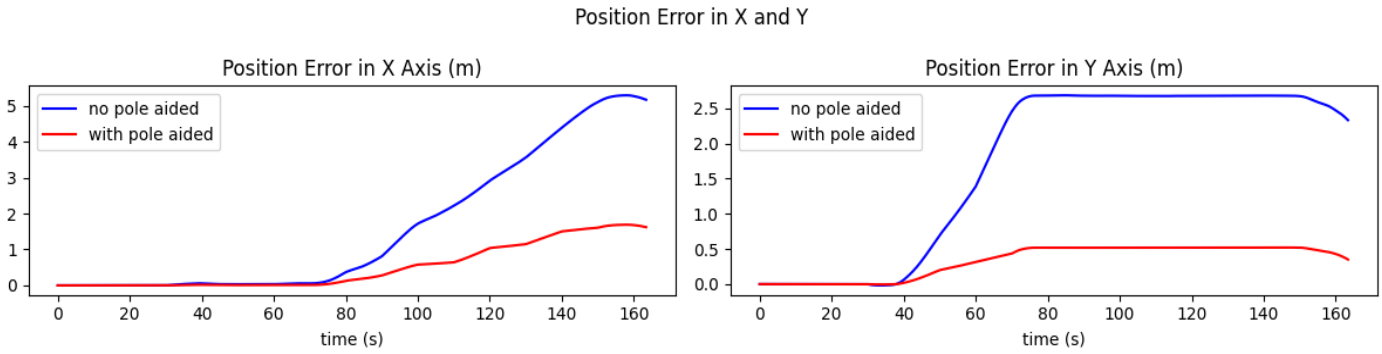

Figure 10 *Translation error comparison between IMU/wheel encoder fusion assisted by pole registration and no assistance.*

Table 1 Relative positioning error comparison

	x-mean (m)	x-max (m)	y-mean (m)	y-max (m)
No pole aided	1.53	5.37	2.04	2.78
Pole aided	0.57	1.64	0.45	0.55

From table 1, it is found that compared to IMU/wheel encoder fusion, the fusion scheme assisted by pole registration is able to reduce the mean translation error in x direction from 1.53m to 0.57m, and mean translation error in y direction from 2.04m to 0.45m. The max translation error in x direction can be reduced from 5.37m to 1.64m, the max translation error in y direction can be reduced from 2.78m to 0.55m.

Secondly, the IMU/wheel encoder/RTK fusion assisted by pole registration and that of no assistance were compared, although the valid RTK fix solution is rare under urban canyon and tree-lined roads. In the evaluation, the distance of adjacent fixed RTK solutions are set 500 meters manually to simulate the long GNSS degraded environment. The evaluation result are given in curve comparison in figure 11 and quantitative comparison in Table 2. In figure 11, curves' color of translation error were same as former. Blue line represent the translation error of IMU/wheel encoder/RTK fusion and red line represent translation error of IMU/wheel encoder/RTK fusion assisted by pole registration. It is shown that the translation error of scheme assisted by pole registration is much less than that of no pole assistance scheme.

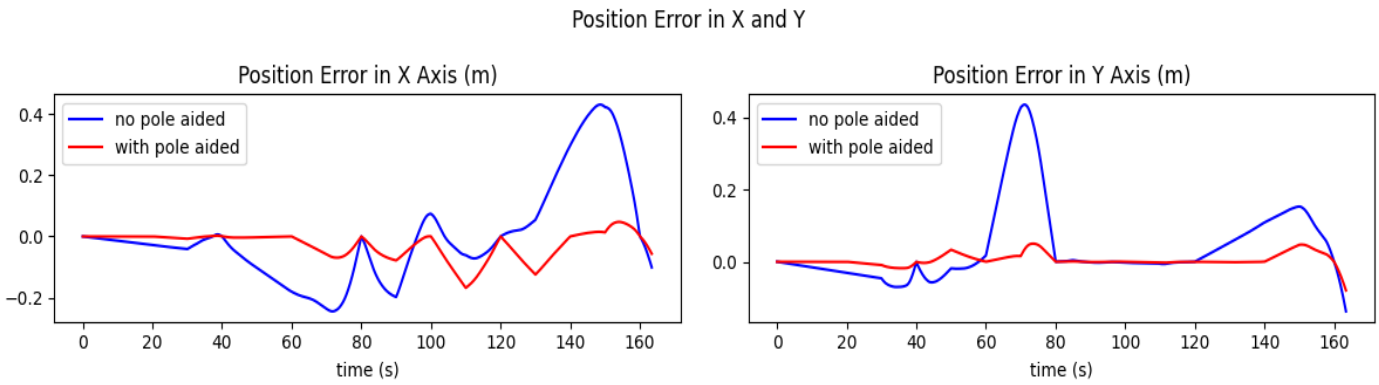

Figure 11 *Translation error comparison between IMU/Wheel Encoder/RTK fusion assisted by pole registration and no assistance.*

Table 2 Absolute positioning error comparison

	x-mean (m)	x-max (m)	y-mean (m)	y-max (m)
No pole aided	0.17	0.42	0.16	0.42
Pole aided	0.08	-0.18	0.05	-0.08

It is shown that when there is no fixed RTK solution in 500 meters distance, compared to IMU/wheel encoder/RTK fusion, the fusion scheme assisted by pole registration is able to reduce the mean translation error in x direction from 0.17m to 0.08m, and mean translation error in y direction from 0.16m to 0.05m. More significantly, the max translation error in x direction can be reduced from 0.42m to 0.18m, the max translation error in y direction can be reduced from 0.42m to 0.08m. The translation error of IMU/wheel encoder/RTK fusion assisted by pole registration can meet the requirement of HD Map's making.

VI. CONCLUSION

GNSS-RTK solutions usually cannot be fixed under urban canyon or tree-lined roads, which makes the HD Map precision cannot meet the requirement of autonomous driving. The paper proposed an effective registration algorithm between pole-like objects detected by point-cloud and that detected by semantic segmentation on image plane algorithm, the relative pose estimation between two distant states are built by pole registration, which reduced the relative pose drift dramatically. Furthermore, a multi-sensor-fusion consists of relative pose estimation from IMU/wheel encoder fusion and relative pose estimation from pole-registration and GNSS-RTK direct observation are built, max absolution positioning error reduced from 0.42 m to 0.18 m under 500 meter's GNSS denied environment.

REFERENCES

[1] Doer, C., Henzler, M., Messner, H. & Trommer, G. F., (2020) HD Map Generation from Vehicle Fleet Data for Highly Automated Driving on Highways. 2020 IEEE Intelligent Vehicles Symposium (IV), 2014-2020, doi:.1109/IV47402.2020.9304781.
[2] Duckworth, G. L., & Baranoski, E. J., (2007) Navigation in GNSS-denied environments: Signals of opportunity and beacons. Proceedings of the NATO Research and Technology Organization (RTO) Sensors and Technology Panel (SET) Symposium. 1-14.
[3] Litman, Y., McGann, D., Dexheimer, Eric., & Kaess, M., (2022) GPS-Denied Global Visual-Inertial Ground Vehicle State Estimation via Image Registration. 2022 International Conference on Robotics and Automation (ICRA), 8178–8184, https://doi.org/10.1109/ICRA46639.2022.9812364
[4] Fischer, P., Azimi, S. M., Roschlaub, R., & Krauß, T., (2018) Towards HD maps from aerial imagery: Robust lane marking segmentation using country-scale imagery. ISPRS International Journal of Geo-Information, 7(12), 458.
[5] Groves, P. D., (2013) Principles of GNSS, inertial, and multisensor integrated navigation systems (2nd Edition). Artech House.
[6] Braasch M. S., (2017) Multipath. In: Teunissen PJG, Montenbruck O, eds. Springer Handbook of Global Navigation Satellite Systems.London, UK: Springer, 443-468.
[7] Jiang, Z., & Groves, P. D., (2014) NLOS GPS signal detection using a dualpolarization antenna. GPS Solutions. 18(1):15-26.
[8] Groves, P. D., & Adjrad, M. (2019). Performance assessment of 3D-mapping–aided GNSS part 1: Algorithms, user equipment, and review. Navigation, 66(2), 341-362.

[9] Adjrad, M., Groves, P. D., Quick, J. C., & Ellul, C. (2019). Performance assessment of 3D-mapping-aided GNSS part 2: Environment and mapping. Navigation, 66(2), 363-383.
[10] Wen, W., Zhang, G., and Hsu, L. T. (2019). Correcting NLOS by 3D LiDAR and building height to improve GNSS single point positioning. NAVIGATION, Journal of the Institute of Navigation, 66(4), 705-718.
[11] Wen, W., & Hsu, L. T. (2022). 3D LiDAR aided GNSS NLOS mitigation in urban canyons. IEEE Transactions on Intelligent Transportation Systems, doi: 10.1109/TITS.2022.3167710.
[12] Wan, G., Yang, X., Cai, R., Li, H., Zhou, Y., Wang, H., & Song, S. (2018). Robust and precise vehicle localization based on multi-sensor fusion in diverse city scenes, 2018 IEEE International Conference on Robotics and Automation (ICRA), 4670-4677.
[13] Alrousan, Q., Alzu'bi, H., Pfeil, A., and Tasky, T., (2020) Autonomous Vehicle Multi-Sensors Localization in Unstructured Environment," SAE Technical Paper 2020-01-1029, 2020, https://doi.org/10.4271/2020-01-1029.
[14] Qin, T., Li, P., and Shen, S. (2018). Vins-mono: A robust and versatile monocular visual-inertial state estimator. IEEE Transactions on Robotics, 34(4), 1004-1020.
[15] Campos, C., Elvira, R., Gómez Rodríguez, J. J., Montiel, J.M.M., & Tardós, J.D. (2021). ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM, IEEE Transactions on Robotics, (37) 1874-1890, doi: 10.1109/TRO.2021.3075644.
[16] Shan, T., Englot, B., Meyers, D., et al. (2020). LIO-SAM: Tightly-coupled lidar inertial odometry via smoothing and mapping. In 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 5135-5142.
[17] Ye, H., Chen, Y., and Liu, M. (2019). Tightly coupled 3D lidar inertial odometry and mapping. In 2019 International Conference on Robotics and Automation (ICRA), 3144-3150.
[18] Tang, J., Chen, Y., Niu, X., Wang, L., Chen, L., Liu, J., ... & Hyyppä, J. (2015). LiDAR scan matching aided inertial navigation system in GNSS-denied environments. Sensors, 15(7), 16710-16728.
[19] Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., ... & Xiao, B. (2020). Deep high-resolution representation learning for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 43(10), 3349-3364.
[20] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778.
[21] Kang, Z., Yang, J., Zhong, R., Wu, Y., Shi, Z., & Lindenbergh, R. (2018). Voxel-based extraction and classification of 3-D pole-like objects from mobile LiDAR point cloud data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(11), 4287-4298.
[22] Liu, R., Wang, P., Yan, Z., Lu, X., Wang, M., Yu, J., ... & Ma, X. (2020). Hierarchical classification of pole-like objects in mobile laser scanning point clouds. The Photogrammetric Record, 35(169), 81-107.
[23] Qi, C. R., Liu, W., Wu, C., Su, H., & Guibas, L. J. (2018). Frustum pointnets for 3d object detection from rgb-d data. In Proceedings of the IEEE conference on computer vision and pattern recognition, 918-927.