

Integration of Landmark Detection and Low-cost Sensors for Vehicle Localization in Challenging Environments

Yu Hsiang Wang
 Electrical Engineering
 National Cheng Kung University
 Tainan, Taiwan
 e-mail: ex4587@gmail.com

Jyh Ching Juang
 Electrical Engineering
 National Cheng Kung University
 Tainan, Taiwan
 e-mail: 8202019@gs.ncku.edu.tw

Muhammad Rony Hidayatullah
 Electrical Engineering
 National Cheng Kung University
 Tainan, Taiwan
 e-mail: mronyh97@gmail.com

Abstract—A seamless vehicle localization capability with high accuracy and integrity is essential for the safe operation of automated vehicles. This study integrates a map-matching based detection scheme and a low-cost Global Navigation Satellite System (GNSS) and Inertial Measurement Unit (IMU) system to enhance the localization capability in a challenging environment. Existing vehicle navigation systems typically use a GNSS/IMU navigation suite to provide position, velocity, and attitude. Such a navigation suite is subject to the error characteristics of the IMU and the operating environment of the GNSS. If the GNSS signals are affected for a long period of time and the quality of the IMU is not well calibrated, erroneous navigation results may occur. It is noted that a challenging environment is featured with some landmarks such as traffic lights. The significant visual feature can be detected robustly by using a deep learning model in a whole day time, which means the availability of the proposed method is better than previous vision-based localization schemes. The paper investigates the fusion of a low-cost GNSS receiver, IMU, vehicle odometer, monocular camera, and an HD map to render seamless navigation. The system is implemented in a vehicle and tested at Taiwan CAR Lab. The effectiveness of the proposed scheme is demonstrated.

Keywords—autonomous vehicle; computer vision; vehicle localization; high definition map; sensor fusion.

I. INTRODUCTION

With the development of intelligent transportation systems, automated driving or self-driving has attracted worldwide attention for its potential in enhancing vehicle safety, improving transportation efficiency, and introducing business opportunities. Precise localization is an essential component of autonomous vehicles. A high-accuracy and high-integrity localization result leads to high-performance path planning, decision-making, and motion control behaviors. Typically, the localization is implemented by using a Global Navigation Satellite System receiver, which results in acceptable performance in the absence of signal obstruction. However, the availability of GNSS-based localization suffers from signal blocking, multipath effect, and atmospheric signal distortions. The integration of the GNSS receiver and Inertial Navigation System (INS) has become an important vehicle navigation suite as GNSS and INS are complementary and can be intelligently fused to render

continuous location and attitude information. The use of GNSS Real Time Kinematic (RTK) technique [1] can further improve accuracy. Another technique is Normal Distribution Transform (NDT) [2] using the lidar to estimate the location by matching the point cloud. Although both of these two methods can achieve centimeter-level accuracy localization, they cost a lot.

To reduce the localization error of a low-cost GNSS/INS, it is required to develop another technique to achieve enough localization accuracy in a dense urban environment. Map-matching based localization systems by using visual features and point cloud are widespread. The concept of map-matching based methods is to detect road elements such as poles, traffic signs, and road markings via perception sensors and find the correspondences of landmarks to help deduce the actual vehicle's position. Map-matching based localization approaches can be generally divided into three categories: (1) The Kalman filter-based localization estimates the vehicle position by matching features on an image with map information. Pink *et al.* [3] proposed to estimate the vehicle's location by using a stereo camera rig to match the visual measurement to a digital feature map. Weiss *et al.* [4] used lidar features associated with precise landmark maps to deduce the vehicle's location based on an Extended Kalman Filter (EKF). (2) Monte-Carlo signal-level localization methods use the raw data to update the state without doing feature extraction. Mattern *et al.* [5] used a coherency value derived from the structure tensor to directly update the image. (3) The last localization approach category is feature extraction based Monte-Carlo localization. Within the Monte-Carlo localization approach, Schindler *et al.* [6] deduced the position by integrating perception information from a monocular camera and laser scanner and associated the landmarks in the high-precision digital map.

Recently, the "High-Definition Map (HD Map)" [7] has become a major research topic to help satisfy the high accuracy demand in Advanced Driver Assistance Systems, as well as the self-driving industry. In contrast with previous digital maps, an HD map includes more precise road geometry, slope, and new features for vehicle localization and perception. In other words, an HD map with highly detailed three-dimensional information can make the vehicle operate more wisely. Apart from using an HD map to improve vehicle localization performance, it can be utilized to enhance detection accuracy by projecting

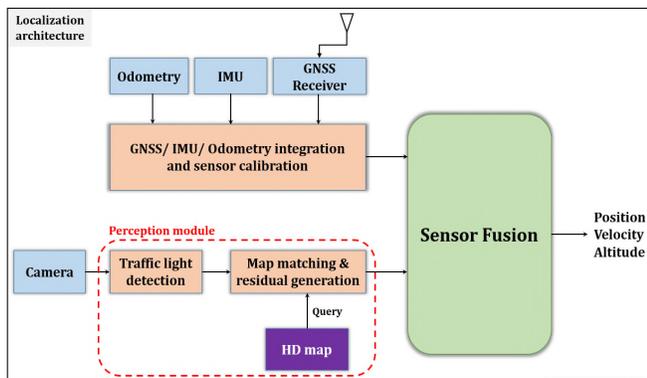


Figure 1. Overview of the self-localization approach.

the information of landmarks stored in the map. Taking advantage of this feature, this paper focuses on integrating an HD map with landmark features, such as traffic lights, to develop a map-matching based localization system. There are several reasons to target traffic lights as the landmark features on an image. First, owing to the detection of road features by using a monocular camera, the quality of the image is sensitive to illumination, exposure, and weather. Traffic lights have significant vision characteristics on an image that can make perception systems get more robust detection results. Second, compared to other road features, traffic lights have long-term stability which is more reliable in terms of its position information on the map. Third, an HD map can constitute prior information for generating region of interest on an image that can improve traffic light recognition and reduce false positives.

Considering the background described above, the main purpose of this paper is to fuse different sensors based on Kalman filter and combine an HD map to overcome the difficulty of vehicle localization inside tunnels. Moreover, in view of the whole system architecture, Region Of Interest (ROI) projection querying from the map is used not only to estimate observations, but also to help recognize traffic lights.

The rest of the paper is structured as follows. In Section II, we describe the idea of our system, which is a vision-aided loosely coupled framework. Section III describes how to obtain correction information by image processing. Section IV goes into the multi-sensor fusion based on an EKF for vehicle localization. Section V demonstrates the experimental results. Finally, we conclude the work in Section VI.

II. SYSTEM OVERVIEW

An overview of the proposed localization architecture is illustrated in Figure 1. The main idea of the approach to determine the vehicle position is based on the Kalman Filter. The integration of GNSS and INS has been well investigated in the literature. Different integration strategies have been exploited and analyzed with different levels of IMU [8][9]. According to the type of operating systems and applications, a specific strategy can be

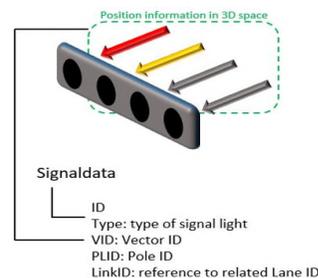


Figure 2. Traffic light information in an HD map.

chosen. Due to the simplicity of implementation and robustness, the loosely coupled integration has been chosen in this paper. The robustness lies in the fact that if one of the systems fails, navigation can still be provided by another sensor [10]. To enhance the capacity and reliability of the localization system for a complex environment, a visual feature-aided method is added to the loosely coupled integration.

The whole framework can be regarded as having two phases. One phase uses the common GPS/INS/Odometry integration strategy. The positions and velocities of the vehicle derived by GNSS signal processing are merged as updates of the INS estimates via a Kalman filter. The other phase uses an observation model for providing correction information by associating the traffic light measurement from the monocular camera with the corresponding information in the high-definition map. Therefore, the system exploits Traffic Lights (TLs) in the testing field as visual features. To detect traffic lights robustly against similar objects, such as a backlight of a vehicle or an external light source, the usage of the HD map as the prior knowledge to generate ROIs can not only drastically reduce false positive detection results, but also can be used to identify the status of the traffic light. In this context, we focus on elaborating on the development of the map-matching based scheme, which improves the capability of the detector and the accuracy of low-cost devices for navigation.

III. IMAGE PROCESSING

To generate correction information for updating the status of the vehicle's location, the detected landmark features should be matched to the corresponding information in the HD map. In Section III, the integration process of the monocular camera and the HD map are introduced, as follows.

A. Landmark Projection

Establishing an explicit coordinate system is a very important step to a multi-sensor fusion methodology. In this paper, the process of mapping the information from the HD map to the image plane is achieved by the coordinate transformation. The coordinate conversions include three different kinds of coordinate systems which are world coordinate, vehicle coordinate, and sensor coordinate, respectively. The HD map provides location information with WGS-84 coordinate, while the

navigation frame (n-frame) is the North East Down (NED) Cartesian coordinate. Hence, the conversion between WGS-84 and NED local coordinate system can be written as the following formula (1):

$$P_{NED} = R^T (P_{ECEF} - P_{Ref}) \quad (1)$$

where P_{NED} is a 3D position in a NED coordinate system, converted from P_{ECEF} ECEF position with respect to reference ECEF position P_{Ref} . Each signal data of a traffic light in the HD map consists of a pole and light bulbs presented in vectors aligned with the center position of bulbs, as shown in Figure 2. For practical utilization of landmark features, mainly front orientated traffic lights will be projected in an image. The origin of the camera frame is the lens optical center and the optical axis is the z-axis of the camera frame. According to the current position of the vehicle, the monocular camera can extract 3D information of traffic lights from the HD map, and only the light markers which are contained in the line-of-sight of the camera will be mapped to the image plane via coordinate transformations (2).

$$P^c = R_n^c P^n + T_n^c \quad (2)$$

where P^n represent the traffic light position in the navigation frame, and derived the point of the traffic light in the camera frame P^c by rotation and transformation from the navigation frame to the camera frame. Conversion from navigation frame to vehicle frame (body frame) is according to the position and the orientation of the vehicle. Because the monocular is mounted on the vehicle, thus, the relationship between the body frame and camera frame is right the extrinsic parameters of the camera model, which can be obtained by offline camera calibration. By finishing the camera calibration, both intrinsic and extrinsic parameters can be known. Eventually, the re-projection process is done by being derived with intrinsic parameters. The pixel coordinates are denoted as u and v . The two-dimensional image plane can be denoted as (3)

$$u = \frac{f_x \cdot x_c}{z_c} + c_x, v = \frac{f_y \cdot y_c}{z_c} + c_y \quad (3)$$

B. Traffic Light Recognition

Vision-based traffic light recognition methods have been widely investigated by using feature-learning models [11], which can detect targets within ROIs. However, only a vision-based method has a lot of challenges such as the influence of weather conditions, varying illumination, viewpoints, and so on. Owing to the rapid development of deep neural networks, a deep model can classify and localize objects accurately.

In 2019, H. Law and J. Deng proposed the CornerNet [12] which detects an object with a pair of key points. Due to objects detected as paired keypoints, the design of using anchor boxes as single-stage detectors was modified. After the novel approach of keypoints,



Figure 3. Model inference at a different time of a day.



Figure 4. (a) The projection of traffic lights' center points. (b) ROI generation according to the projection from the map.

CenterNet [13], detecting objects as axis-aligned boxes, uses keypoint estimation to search center points of objects and then optimizes the objective function to regress to all other object properties, such as object size and location. Owing to its faster inference time and higher accuracy, it is applied to recognize TLs in different illumination of a day time shown in Figure 3.

Although using a deep network to detect TLs, false-positives might be generated because of external light sources or the influence of illumination variation. Various approaches extending localization and map information as prior knowledge for traffic light recognition have been proposed [14]. Therefore, we adopt CenterNet to detect targets that should be verified as true TLs in ROIs generated by the HD map. It shows that utilizing the map as prior information can dramatically reduce noise that disturbs the model shown in Figure 4(b).

C. Data Association

When the deep learning model detects lighted marks in certain sections, there may be more than one traffic light in an image. To obtain the re-projection errors as correction information, the detection results from the deep neural network and re-projection of corresponding TLs from the map should be matched correctly. The

distance $y(\tilde{d}_k, \hat{d}_k)$ between the detection \tilde{d}_k^i and the projection \hat{d}_k^j of the traffic light at epoch k is hereby



Figure 5. Data association between detection results and projection (gray virtual box).

formulated as a combination of the Mahalanobis distance [15] according to their positions (4):

$$y(\tilde{d}_k, \hat{d}_k) = \alpha \cdot \delta d_k^{iT} S_k^{-1} \delta d_k^i + (1 - \alpha) \cdot \frac{\omega_d^2}{\sigma_d^2} \quad (4)$$

where δd_k^i is Euclidean distance of $\tilde{d}_k^i - \hat{d}_k^j$, S_k is the covariance of δd_k^i , ω_d is the width of the bounding box of the detected TL, while σ_d is the width of projection, and the additional factor α is used to weight the impact of the position and width of a TL. If the Mahalanobis distance of each detected TL is lower than the threshold, the re-projection errors between detection and projection results can be regarded as residuals for updating the status of the vehicle shown in Figure 5. Finally, the correction information can be utilized by the visual measurement model for EKF, which will be introduced in Section IV.

IV. SENSOR FUSION FOR LOCALIZATION

In our work, the Extended Kalman filter has been chosen to accomplish the self-localization for the vehicle. In state formulations, the state equation at each epoch is propagated with the INS dynamic model and updates each state with measurements from multi-sensors.

A. State Model

An INS with fifteen states was developed and the complete state is denoted as:

$$X = [p^n \ v^n \ \rho \ b_a \ b_g]^T \quad (5)$$

where p^n is the vehicle position in the navigation frame, v^n is the velocity of the vehicle, ρ is attitude including roll, pitch, and yaw angle, b_a is the bias of the accelerometer, b_g is the bias of the gyroscope. The nominal-state kinematics corresponding to the system without noises or perturbations can be denoted as (6)

$$\begin{aligned} \dot{p}^n &= v^n \\ \dot{v}^n &= R_b^n \cdot f_{ib}^b - (2\Omega_{ie}^n + \Omega_{en}^n) \cdot v^n + g^n \\ \dot{R}_b^n &= R_b^n (\Omega_{ib}^b - \Omega_{ie}^b) \end{aligned} \quad (6)$$

where f_{ib}^b represents the acceleration of the vehicle in b-frame, R_b^n is the rotation matrix from body frame to navigation frame, $(2\Omega_{ie}^n + \Omega_{en}^n) \cdot v^n$ represents the Coriolis acceleration and g^n is the gravitational acceleration given by gravity model. Ω_{ib}^b is the skew-symmetric matrix of the angular velocity in b-frame. Ω_{ie}^b is the skew-symmetric matrix of the angular velocity of the Earth's rotation. The state model is formulated in discrete-time corresponding to the real system. The prediction stage consists of predicting the state using knowledge of the previous epoch, as (7)

$$X_{k+1}^- = A_{k+1} X_k^+ + \Gamma_{k+1} w_k \quad (7)$$

where A_{k+1} is the state transition matrix at epoch k . w_k is the process noise transformed by Γ_{k+1} to body frame.

$$P_{k+1}^- = \Phi_{k+1} P_k^+ \Phi_{k+1}^T + Q_k \quad (8)$$

where the transformation matrix Φ_{k+1} can be approximately as $\Phi_{k+1} = I + F \cdot dt$ by using the first order of the Taylor series.

B. GNSS/Odometry Measurement Model

For the common loosely-coupled integration, the GNSS measurement model can be calculated by:

$$z_k = \begin{bmatrix} r_{GNSS}^n - r_{INS}^n \\ v_{GNSS}^n - v_{INS}^n \end{bmatrix} \quad (9)$$

where r_{GNSS}^n and v_{GNSS}^n are the position and velocity derived by GNSS in n-frame. r_{INS}^n and v_{INS}^n are position and velocity that derived by INS mechanism in n-frame. The observation matrix for updating the GNSS measurement can be derived as (10)

$$H_k = \begin{bmatrix} I_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 9} \\ 0_{3 \times 3} & I_{3 \times 3} & 0_{3 \times 9} \end{bmatrix} \quad (10)$$

The Odometry measurement is written as follows:

$$z_k = [speed_{Odometry}^n - speed_{INS}^n] \quad (11)$$

$$speed_{Odometry}^n = \frac{1}{2} \cdot r_{wheel} \cdot (\omega_{rl} + \omega_{rr}) \quad (12)$$

$$speed_{INS}^n = \sqrt{v_n^2 + v_e^2 + v_d^2} \quad (13)$$

where ω_{rl} and ω_{rr} are the angular velocities of the left and right wheels, respectively. r_{wheel} is the radius of the wheels and the observation model can be written as (14)

$$H_k = \begin{bmatrix} 0_{1 \times 3} & \frac{v_n}{speed_{INS}^n} & \frac{v_e}{speed_{INS}^n} & \frac{v_d}{speed_{INS}^n} & 0_{1 \times 9} \end{bmatrix} \quad (14)$$

C. Visual Measurement Model

The re-projection result of the landmark, changing with the camera's pose, can be described as the observation model denoted as (15).

$$\hat{d}_k = h(\hat{X}_k^+) \quad (15)$$

where h is the measurement function that extracts the nearby traffic lights to convert to the measurement at k epoch. The residual between the visual measurement and the predicted value can be given in (16).

$$\delta d_k = \tilde{d}_k - \hat{d}_k = H_k \delta X_k + v_k \quad (16)$$

where $H_k = \left. \frac{\partial h}{\partial X} \right|_{X=\hat{x}_k^+}$ is the Jacobians of the estimated

measurement with respect to the state vector and v_k is the observation noise that correlated with R_k . Based on the knowledge of 3D reconstruction optimization algorithm, Bundle adjustment [16], the relationship between the error on image and the camera pose can be derived by minimizing the re-projection error described as (17)

$$\delta d_k^* = \arg \min_{\delta d_k} \frac{1}{2} \sum_{i=1}^n \left\| \tilde{d}_k^i - \hat{d}_k^i \right\| \quad (17)$$

$$\hat{d}_k^i = \frac{1}{s} \cdot K \cdot [R | T]_n^c \cdot P^n \quad (18)$$

where s is the scale factor. K is the intrinsic parameters of the camera model, obtained by camera calibration [17], and P^n is the landmark position in the map. By linearizing the error with the first order of the Taylor series, the corresponding Jacobian matrix can be derived as (19)

$$H_k = K_c D C \cdot \begin{bmatrix} -I_{3 \times 3} & 0_{3 \times 3} & [\hat{P}^c \times] & 0_{3 \times 6} \end{bmatrix} \quad (19)$$

where

$$K_c = \begin{bmatrix} f_y & 0 \\ 0 & f_x \end{bmatrix} \quad (20)$$

$$D = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} \quad (21)$$

where

$$d_{11} = 1 + k_1 r^2 + k_2 r^4 + 2p_1 x_c' + 4p_2 x_c'$$

$$d_{12} = 2p_1 y_c'$$

$$d_{21} = 2p_1 y_c'$$

$$d_{22} = 1 + k_1 r^2 + k_2 r^4 + 2p_2 x_c' + 4p_1 y_c'$$

C is the partial derivative of the re-projection error with respect to the position in camera coordinate:

$$C = (1/(\hat{z}_c)^2) \begin{bmatrix} \hat{z}_c & 0 & -\hat{x}_c \\ 0 & \hat{z}_c & -\hat{y}_c \end{bmatrix} \quad (22)$$

$[\hat{P}^c \times]$ is the skew-symmetry matrix of the landmark position in camera coordinate k .

Once there exist traffic lights in an image, the residuals calculated by the difference between the detection result and the re-projection mentioned in Section III can be used to correct the localization error. The Kalman gain at k epoch can be calculated as (23)

$$K_k = P_{k+1}^- H_k^T * [H_k P_{k+1}^- H_k^T + R_k]^{-1} \quad (23)$$

Therefore, the state vector can be updated (24).

$$X_{k+1}^+ = X_{k+1}^- + K_k \cdot \delta d_k \quad (24)$$

The Kalman filter calculates the updated covariance P_{k+1}^- after getting the state estimation, which will be used in the next time step.

$$P_{k+1}^+ = [I - K_k H_k] P_{k+1}^- \quad (25)$$

There might be multiple traffic lights in an image at the same time. To get the measurements that can truly correct the localization errors, the two closest TLs, which are not affected significantly by the geometry distortion of the camera, are chosen as visual measurements to update the state vector. Each measurement is used to update the state estimation and re-calculate the Kalman gain once again. By selecting the specific traffic lights as measurements, the observation values are more reliable and can avoid the bad correction while detection results mismatch the corresponding information in the HD map.

V. EXPERIMENTAL RESULTS

In this section, the evaluation of the proposed method was verified at the Taiwan CAR Lab test facility where various environmental complexities simulate street conditions in Taiwan. The HD map in the testing field was produced by the High Definition Maps Research Center at National Cheng Kung University. In our work, the point $22.99665875N^\circ$, $120.222584889E^\circ$ is set as the origin of the local tangent plane coordinate system and we chose the driving route including different types of traffic lights and a tunnel. Let the vehicle drive counterclockwise to test the capacity of the map-matching based localization system.

A. System configuration

As shown in Figure 6, the level of centimeter localization result from NDT, with lidar VLP 16, is set as the reference. The locating signals are received by the antenna to the GNSS receiver. Except for GMSL camera connected to Nvidia Drive PX2, the other on-board sensors are connected to Industrial PC. The monocular camera is rigidly coupled to the vehicle. The camera optical axis is aligned to the driving direction that can collect the front view image data. The specification of the sensors is shown in Table I. Based on Robot Operating System (ROS), we integrated these two computing core platforms - IPC and Nvidia Drive PX2- to develop our approach.

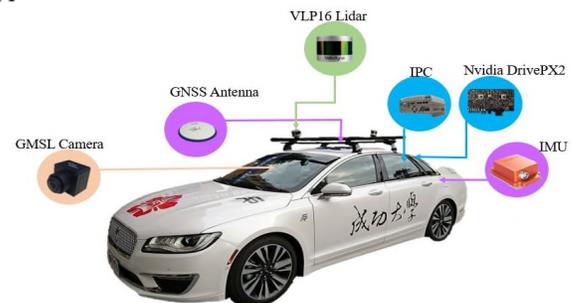


Figure 6. Lincon MKZ with high-performance computing cores and sensors.

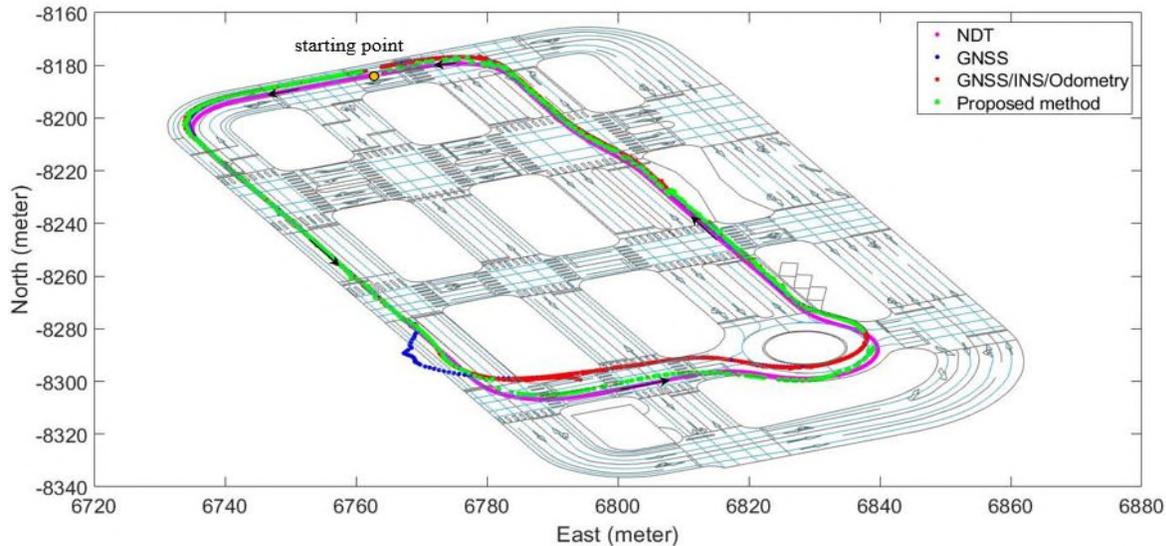


Figure 7. The overall localization results show on the high-definition map. The x - y axis represents the local tangent plane coordinates: east and north respectively. Violet: NDT as the reference. Blue: Only GNSS localization. Red: Integrated GNSS with INS. Green: Corrected localization error by fusing with landmark features in an image and the corresponding information in the HD map.

TABLE I. TABLE TYPE STYLES

Sensor	Model	Type	Parameter
GNSS receiver	HDM2024 EVK	Positioning Accuracy	2.5 m
		Velocity Accuracy	0.1 m/s
		Time Accuracy	25 ns
		Velocity Limit	515 m/s
Camera	GMSL	Resolution	1928x1208
		Optical format	1/2.7 inch
		Field Of View	60 degree

B. Localization results

In this experiment, the maximum localization error happened when the vehicle drove into the tunnel because of the signal outage. The condition of positioning gradually recovered while leaving the tunnel. As the situation mentioned above, the localization system can not only rely on GNSS; otherwise, it will cause fatal accidents for the vehicle. To reduce the localization error, the inertial navigation system can be used to enhance the consistency of updating the position.

Although the fusion of GPS, IMU, and Odometry can provide more reliable navigation solutions that allow the vehicle to drive within the lane, the vehicle still drove against the traffic shown in Figure 7. Owing to the characteristics of the integration of these sensors based on an EKF, the localization estimation had more confidence in the solution from IMU when the vehicle lost signal from GNSS. The drawback of error accumulations from IMU for a period led the vehicle to drive on the wrong side of the road. The proposed method can obtain visual measurements by using the monocular camera to overcome the problems of the conventional GNSS/INS integration approach.

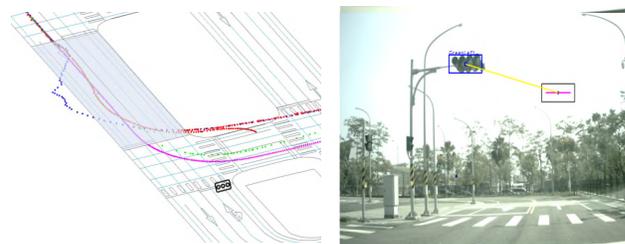
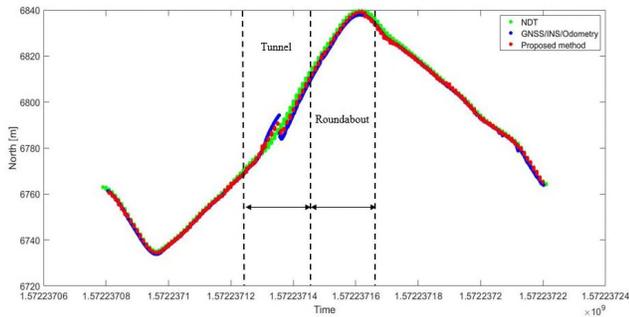


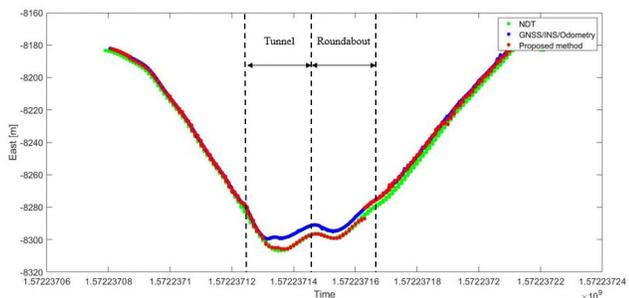
Figure 8. (a) The mask represents the tunnel section in the HD map. (b) The residual between the detection result and the re-projection of the traffic light indicated in (a).

Since the camera is sensitive to the light intensity variation, it was hard to detect the traffic light when the vehicle departed from the tunnel that changed the view from low illumination to high exposure. Compared to the conventional vision learning-based methods, the powerful deep learning model can get more robust detection results against the quality of the image. Therefore, the challenge of varying illumination can be overcome to obtain the correction information for adjusting the state of the vehicle effectively.

Figure 8(a) shows that it is a big challenge for the localization system because of the GNSS/INS localization performance degradation when the vehicle was going to drastically turn left. The residual between the visual measurement detected by the deep model and the re-projection shown in Figure 8(b) can provide significant correction information to help correct the localization error, especially the outcome in the east direction in Figure 9(b). The localization errors in North and East direction are calculated as Table II, which shows that the improvement of the localization accuracy is greater than fifty percent.



(a)



(b)

Figure 9. Localization in each direction (a) Localization in North direction. (b) Localization in East direction.

TABLE II. TABLE TYPE STYLES

	GNSS	GNSS/IMU/Odometry	Proposed method
North error [m]	4.22	4.10	1.85
East error [m]	2.65	1.99	1.32

In addition, the proposed method can provide the vehicle's attitude information. Figure 10 refers to the heading angle comparison. It can be seen that the heading drifted due to GNSS signal-degraded when the vehicle passed through the tunnel. Owing to the loosely coupled integration using IMU, the heading angle is corrected. In particular, the correction information generated by the camera and HD map can be used to dramatically improve the heading.

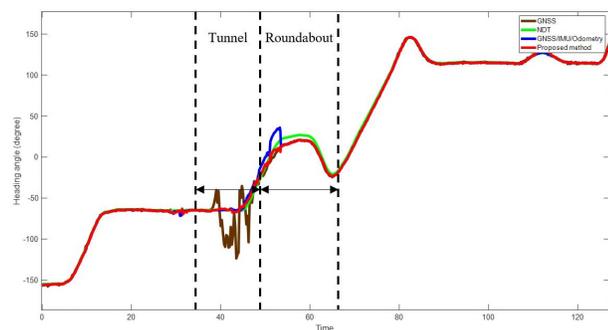


Figure 10. Heading angle (degree) comparison.

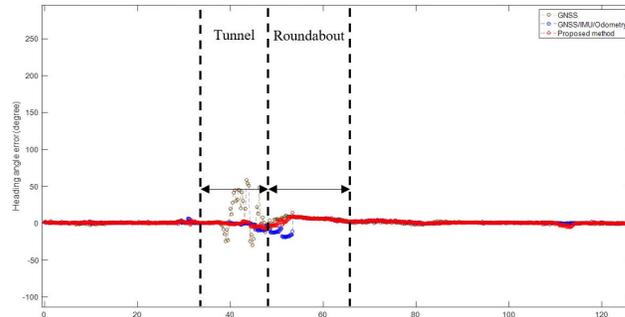


Figure 11. Heading angle (degree) error.

Further comparing to the ground truth, the heading errors are depicted in Figure 11. The heading errors can be averaged below 5 degrees.

VI. CONCLUSION AND FUTURE WORK

The map-matching based localization scheme demonstrated in this work shows that the integration of visual features and an HD map can dramatically improve a low-cost GNSS, IMU and Odometry integration. This paper is aimed to solve the problem when the GNSS/INS/Odometry localization system fails to locate a reliable position under some circumstances, especially in a tunnel. The experimental results show that even the localization drifts caused by the GNSS signal outage and IMU error accumulation, as long as landmark features can be extracted by the monocular camera, then the residuals computed by the re-projection error in an image can effectively correct the positioning error. Moreover, the proposed method can provide a more accurate vehicle heading angle.

Although the availability of the proposed method is subject to intersections, the strong visual features of traffic lights can make it up in a challenging environment. The integration of traffic light recognition and an HD map can not only reduce the false-positive detection results, but also provide useful correction information for the system update. Besides, the camera calibration is a very important prerequisite for the visual observation model, because the change of each direction or orientation will significantly affect the projection result. Therefore, further researches may pay attention to on-line camera calibration to avoid long-term drifting of the camera parameters. To overcome the limitation that TLs can be only detected at intersections, other road features in an image should be considered such as lane lines, traffic signs and pole-like objects, which can provide more useful information for vehicle localization.

In conclusion, the map-matching based system using traffic lights as visual features can significantly improve an integration strategy with low-cost devices for vehicle localization. Moreover, the availability of a vision-aided loosely coupled framework can be enhanced and it can be improved by solving the problems mentioned above.

ACKNOWLEDGMENT

This research was funded by the Ministry of Science and Technology, Taiwan, grant number MOST 108-2218-E-006-052.

REFERENCES

- [1] R. J. Keller, M. E. Nichols, and A. F. Lange, "Methods and apparatus for precision agriculture operations utilizing real time kinematic global positioning system systems," ed: Google Patents, 2001.
- [2] P. Biber and W. Straßer, "The normal distributions transform: A new approach to laser scan matching," in *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453)*, 2003, vol. 3: IEEE, pp. 2743-2748.
- [3] O. Pink, F. Moosmann, and A. Bachmann, "Visual features for vehicle localization and ego-motion estimation," in *2009 IEEE Intelligent Vehicles Symposium*, 2009: IEEE, pp. 254-260.
- [4] T. Weiss, N. Kaempchen, and K. Dietmayer, "Precise ego-localization in urban areas using laser scanner and high accuracy feature maps," in *IEEE Proceedings. Intelligent Vehicles Symposium*, 2005: IEEE, pp. 284-289.
- [5] N. Mattern and G. Wanielik, "Camera-based vehicle localization at intersections using detailed digital maps," in *IEEE/ION Position, Location and Navigation Symposium*, 2010: IEEE, pp. 1100-1107.
- [6] A. Schindler, "Vehicle self-localization with high-precision digital maps," in *2013 IEEE Intelligent Vehicles Symposium Workshops (IV Workshops)*, 2013: IEEE, pp. 134-139.
- [7] F. Poggenhans *et al.*, "Lanelet2: A high-definition map framework for the future of automated driving," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018: IEEE, pp. 1672-1679.
- [8] A. Solimeno, "Low-cost INS/GPS data fusion with extended Kalman filter for airborne applications," *Masters of Science, Universidade Technica de Lisboa*, 2007.
- [9] M. G. Petovello, *Real-time integration of a tactical-grade IMU and GPS for high-accuracy positioning and navigation*. Citeseer, 2003.
- [10] S. Godha, "Performance evaluation of low cost MEMS-based IMU integrated with GPS for land vehicle navigation application," *UCGE report*, no. 20239, 2006.
- [11] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1484-1497, 2012.
- [12] H. Law and J. Deng, "Cornernet: Detecting objects as paired keypoints," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 734-750.
- [13] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," *arXiv preprint arXiv:1904.07850*, 2019.
- [14] M. Hirabayashi, A. Sujiwo, A. Monrroy, S. Kato, and M. Edahiro, "Traffic light recognition using high-definition map features," *Robotics and Autonomous Systems*, vol. 111, pp. 62-72, 2019.
- [15] R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, "The mahalanobis distance," *Chemometrics and intelligent laboratory systems*, vol. 50, no. 1, pp. 1-18, 2000.
- [16] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment—a modern synthesis," in

International workshop on vision algorithms, 1999: Springer, pp. 298-372.

- [17] A. Dhall, K. Chelani, V. Radhakrishnan, and K. M. Krishna, "LiDAR-camera calibration using 3D-3D point correspondences," *arXiv preprint arXiv:1705.09785*, 2017.