Augmenting Vehicle Localization Accuracy with Cameras and 3D Road Infrastructure Database

Lijun Wei(✉), Bahman Soheilian, and Valérie Gouet-Brunet

IGN, SRIG, MATIS, Université Paris-Est, 73 Avenue de Paris,
94160 Saint Mandé, France
{lijun.wei,bahman.soheilian,valerie.gouet}@ign.fr

Abstract. Accurate and continuous vehicle localization in urban environments has been an important research problem in recent years. In this paper, we propose a landmark based localization method using road signs and road markings. The principle is to associate the online detections from onboard cameras with the landmarks in a pre-generated road infrastructure database, then to adjust the raw vehicle pose predicted by the inertial sensors. This method was evaluated with data sequences acquired in urban streets. The results prove the contribution of road signs and road markings for reducing the trajectory drift as absolute control points.

Keywords: Vehicle localization · Road infrastructure database · Road signs · Road markings

1 Introduction

To compensate the low performance of GPS receiver in dense urban environments caused by multi-path or building occlusions, dead-reckoning sensors like wheel encoder, inertial sensors, or visual odometry method have been integrated to continuously predict the vehicle movement. A main problem of the dead-reckoning methods is the pose error accumulation from point to point, thus a lot of methods have been proposed to alleviate the trajectory drift. *Personal Navigation Devices* use classic map matching method [1] to associate the vehicle location with a digital map of road networks. As the road network map usually well represent the topological relationship between different road segments, but lack of geometric accuracy, some other methods were proposed to generate an enhanced digital map with visual landmarks from onboard perception sensors. Visual landmarks are those static and recognizable objects in the environment. Several systems treat the reconstructed 3D points as landmarks [2][3][4]: interest points and descriptors (SIFT, SURF, HoG, etc.) are detected and extracted from multiple images and reconstructed into 3D point cloud by structure-from-motion with bundle adjustment. If the image sequence is already geo-referenced

L. Agapito et al. (Eds.): ECCV 2014 Workshops, Part I, LNCS 8925, pp. 194–208, 2015.
DOI: 10.1007/978-3-319-16178-5_13

by a localization device, each 3D point in the database is associated with its absolute 3D coordinates, and its 2D locations and visual appearance (descriptors) in the corresponding 2D images. The on-line localization step is then to associate the sensor perception with the landmarks in database, and to recover the current vehicle pose by n-point Direct Linear Transformation (DLT) minimization with RANSAC.

Instead of using the 3D points directly, in this work, we propose to use more semantic and more robust landmarks: a database of 3D road infrastructures, i.e., road signs and road markings, which was automatically generated from georeferenced image sequences, as shown in Fig. 1. Compared with image points, the advantage of using road infrastructure objects is threefold: 1) volume of storage and matching: since the 3D point cloud contains millions of 3D points and corresponding image features, it requires large space for data storage and long time to access the sub point-clouds for landmarks association. As there are fewer road infrastructure objects than the sparse 3D points, it requires less volume for data storage and matching; 2) precision and robustness of landmarks: as the visual appearance of some image points might change during the day or in different seasons (e.g., trees), how to maintain an up-to-date point database is still an ongoing research. Visual landmarks of road infrastructures are more robust, static and precise in urban environments than the sparse points, and the road sign and road marking detection/reconstruction algorithms used can achieve sub-decimeter accuracy as reported in [5] [6]; 3) matching constraint: association of image points is done in multi-dimensional descriptor space and under multi-view geometric constraint, while road infrastructures are semantic visual landmarks with known visual appearance and geometric attributes; the matching step can be based on both geometric and semantic attributes to make it efficient.

Fig. 1. Projection of 3D road sign and road marking landmarks on an image frame with raw camera pose (Left bottom: camera's field of view shown by blue triangle; middle bottom: zoomed view of the projected 3D road sign on image frame)

The most similar work to our study might be [7], in which a camera is used to detect the road markings and laser scanners are used to detect all the distinctive objects (traffic signs, trees, etc.). However, the explicit type and elevation information of the distinctive objects are not known. In [8], a map of curbs and road markings is generated from a stereo pair and used for vehicle localization in rural area. In [9], the authors also mentioned their road object map consisting of manually labeled and reconstructed static road objects like road crossings and road signs from a stereo pair. Their stored map objects are used for yielding an AR system by overlaying the objects on camera images. We use multi-cameras to automatically detect, recognize and reconstruct both the road markings and road signs. These map objects are then stored and used to improve the localization quality, especially in urban environments where GNSS performance is more challenging and the road occlusion is more frequent due to pedestrians and other vehicles. We currently assume that a rough initial vehicle position is provided by a GPS receiver at the beginning of a sequence.

In the remaining of this paper, we firstly introduce the method for generating a 3D road infrastructure database in Sec. 2; then, we present the localization method with 3D road infrastructure landmarks in Sec. 3; finally, some experimental results and discussions are respectively presented in Sec. 4 and Sec. 5.

2 Generation of a 3D Road Infrastructure Database

Road infrastructures include sidewalks, pedestrian crossings, road signs, traffic lights, etc. An accurate and up-to-date 3D database of road infrastructures is not only useful for infrastructure management and maintenance, it can also contribute to advanced driver assistance, like vehicle self-localization, lane keeping/alarming. An infrastructure database is usually manually surveyed and drawn by engineers with portable GPS, this procedure is time-consuming and expensive. This process can be largely accelerated by using ground mobile mapping system (MMS) [10] [11]. Road infrastructure objects are first automatically detected and identified from the acquired scene videos, and then triangulated into 3D with the vehicle poses from a high-precision geo-referencing device.

Fig. 2. (Left) One of the real images used for database generation; (Right) Reconstructed 3D road signs and road marking strips in the database

We follow the pipeline of road infrastructure database generation as in [12]: road markings, i.e. zebra crossings and dashed lines, are automatically detected and reconstructed from a calibrated front-view stereo pair [5]; and road signs are detected, recognized and reconstructed from a multi-camera system on the roof of the vehicle with a constrained multi-view reconstruction method as in [6]. The generated sign/marking database (Fig. 2) is composed of a list of 3D road signs/markings. Each road landmark is encoded with the following information:

(1) simple geometric shape: a road sign is encoded either as a 3D polygon, a triangle, or a circle (not discussed in this paper); a road marking strip is encoded as a parallelogram.

(2) coordinates of road landmark in sub-decimeter accuracy: as the positions of 2D detections are in sub-pixel precision, absolute coordinates (in Easting-Northing-Elevation format) of the corners of each landmark are in sub-decimeter accuracy. This database is also consistent with other geographic maps and environment models.

(3) type of a landmark: a road sign is encoded as "Indication", "Obligation", "Prohibition" or "Warning" type; a road marking strip is encoded as "zebra crossing", or dashed lines with specific types ("T2", "T3", "T'1" or "T'P").

(4) corresponding 2D images used for reconstruction are also listed.

Due to the occlusion by obstacles, it is possible that some road markings or road signs might not be visible in the image sequences used for data base generation. As the road surface of some urban streets is not covered by any markings, we did not use any model to fit or "interpolate" the missed road marking strips. Instead, we consider each road marking strip as a strip patch, which well defines the road marking plane in front of the vehicle.

3 Vehicle Localization with Road Infrastructure Database

In this section, we present the vehicle localization method using an IMU and the aforementioned road infrastructure database. Vehicle pose is firstly predicted by IMU in Sec. 3.1; then, road markings and signs are respectively detected from onboard cameras and associated with the database objects in Sec. 3.2; in addition to the attribute constraint, the Mahalanobis distance between two corresponding landmarks is discussed in Sec. 3.3.

3.1 Vehicle Pose Prediction

Like other visual landmarks based localization systems, we assume that all the sensors are rigidly installed on the experimental vehicle and well calibrated before the experiments. When initial state of the vehicle is given, the vehicle state can be continuously predicted by dead-reckoning systems. To facilitate the propagation of uncertainties between different vehicle positions, a pose-graph can be constructed [13] by considering each vehicle pose as a vertex and the displacement between two consecutive poses as an edge.

We use inertial sensors to provide accelerations and orientations of the vehicle in this work. Let $X_k = [X, Y, Z, \dot{X}, \dot{Y}, \dot{Z}]_k^T$ be the vehicle state at time k, where (X, Y, Z) are vehicle position and $(\dot{X}, \dot{Y}, \dot{Z})$ are vehicle velocities in navigation frame (Easting-Northing-Elevation). Assume that the vehicle acceleration a_k in navigation frame is constant between time step $(k-1)$ and k, the vehicle state at time k can be predicted by:

$$X_k = F_k X_{k-1} + G_k a_k \tag{1}$$

where $F_k = \begin{bmatrix} I_3 & (\Delta T)_{3\times 3} \\ 0_{3\times 3} & I_3 \end{bmatrix}$, I_3 is a 3×3 identity matrix, $G_k = \begin{bmatrix} (\frac{\Delta T^2}{2})_{3\times 3} \\ (\Delta T)_{3\times 3} \end{bmatrix}$, and $\Delta T = T_k - T_{k-1}$. a_k is the vehicle accelerations in navigation frame given by:

$$a_k = R_k a_k^b = R_k [a_x^b, a_y^b, a_z^b]^T = R(\gamma)R(\beta)R(\alpha)[a_x^b, a_y^b, a_z^b]^T \tag{2}$$

where $a_k^b (\in \mathbf{R}^3)$ is the vehicle accelerations in body frame reported by the IMU sensor, R_k is the vehicle attitude (transformation from vehicle body frame at time k to the navigation frame) represented by the vehicle Euler angles (roll α, pitch β and yaw γ) from gyroscope. Assume that the accelerations of vehicle in body frame are respectively perturbed by independent white noises with variance δa_x, δa_y, δa_z, the covariance matrix of a_k^b is written as $C_k = diag(\delta a_x^2, \delta a_y^2, \delta a_z^2)$. The covariance Qx_k of the vehicle pose at time k can be estimated by linearization of Eq. 1:

$$Qx_k = F_k Qx_{k-1} F_k^T + (G_k R_k) C_k (G_k R_k)^T \tag{3}$$

3.2 Matching Criteria between Two Road Landmarks

Meanwhile, road markings and signs are respectively detected using the same algorithm as in map generation stage, except that during the mapping stage, possible 2D road signs are matched over all image frames, while in localization stage, only the three front looking images captured at the same instant are used.

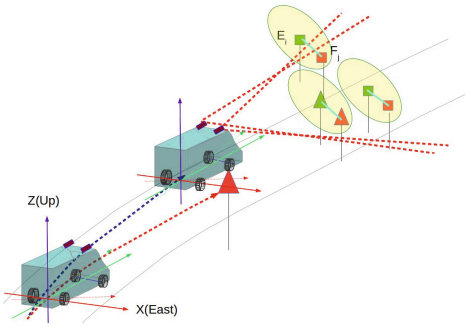


Fig. 3. Vehicle pose correction with visual landmarks (Database landmarks: in red color; online estimation: in green color)

The corresponding 2D detections are fed into the the constrained multi-view reconstruction algorithm as in the mapping stage.

For images at time t with multiple 2D detections S, these 2D detections are first reconstructed into 3D ($\mathcal{E}$) by constrained multi-view reconstruction algorithm, m landmarks can be reconstructed as: $\mathcal{E} = \{E_1, ..., E_i, ..., E_m\}$. Since matching of 2D detections is based on strict geometric and visual appearance constraints, the reconstruction step can help to remove some false positive detections. Then, with the m reconstructed landmarks $\mathcal{E}$, and n reference landmarks in the database: $\mathcal{F} = \{F_1, ..., F_j, ..., F_n\}$, we need to find all the hypotheses to associate each observation E_i with feature F_{j_i} [14]. If there is no matched landmark for E_i, this reconstruction will not be used.

The data association process of road signs and road markings is based on facing direction of landmark plane, Mahalanobis distance, landmarks type and uniqueness constraints, as illustrated in Fig. 3. The Mahalanobis distance is defined by considering both the noises of 3D reconstruction and pose transition process (Eq. 3). The association problem might become ambiguous when the IMU error is very large or the landmarks are densely distributed. If there are multiple candidate landmarks associations, the vehicle state track is split into multiple independent tracks, each within an EKF (Extended Kalman Filter) to correct the vehicle pose. If a GPS measurement is provided, it can be used as a measurement together with the road sign and road marking objects.

Matching Criteria between Two 3D Signs. A 3D road sign observation E_i is matched to a sign F_{j_i} in the database if the following criteria are satisfied:

1) two objects are identified as road signs in the same category;

2) facing directions of the two road sign planes are less than a threshold (set to 40^o in our experiments);

3) Mahalanobis distance between two corresponding road signs is measured by their position difference in the camera frame, this distance should be less than a threshold defined by chi-squared distribution (will be detailed in Sec. 3.3 Eq. 4 to Eq. 6);

4) uniqueness constraint: when multiple road signs are reconstructed from the same image pair, they cannot be associated with the same landmark in the database at the same time.

Matching Criteria between Two Road Marking Objects. In the step of road sign detection/reconstruction, a 2D road sign detection is kept only if the whole sign is seen by the camera for the purpose of type identification. For road markings, this constraint is less strict. Due to the occlusion by obstacles in front of the vehicle, the camera might detect only a portion of a road strip. As the detected strip portion might be at any location inside the corresponding reference strip, a detection uncertainty is added to every strip in the database.

Let the local frame attached to a reference strip is with y axis collinear to the strip, the center of a marking strip reconstructed online might locate along the longitudinal axis, and its lateral position might locate along the lateral axis,

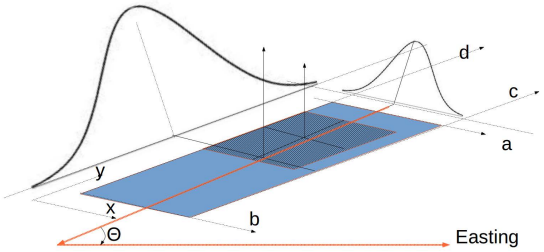


Fig. 4. Matching uncertainty between two strips (blue strip: reference)

as shown in Fig. 4. We set σ_x and σ_y respectively as the local uncertainties of this road strip, $\sigma_x = width/8$ ($width$ is the width of this strip), and σ_y is set as $1m$. The variances of this strip can be transformed into navigation frame with the strip slope θ, as $Q_l = \begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{bmatrix} diag(\sigma_x^2, \sigma_y^2) \begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{bmatrix}^T$.

3.3 Mahalanobis Distance between Corresponding Landmarks

Since our IMU can provide high accurate orientation measurement, vehicle orientation error is not considered in this work and only the noise of acceleration measurements were taken into account in error propagation. The same as other landmarks based localization system with EKF, several blocks of the process is introduced as follows:

1) **Measurement**: if m landmarks are reconstructed at time k, noted as $\mathcal{E}_k = \{E_1, ..., E_i, ..., E_m\}$, $E_i = (\Delta X_i, \Delta Y_i, \Delta Z_i)$ is the 3D position of the center of i^{th} road landmark in current vehicle local frame, Q_{E_i} is the uncertainty of the reconstructed landmark.

2) **Observation**: let R be the attitude of current vehicle state, $F_j(x, y, z)$ be the center of a landmark in the database, the expected 3D position EM_j of the landmark F_j in current vehicle frame can be calculated with the vehicle position X_k and vehicle attitude R, as:

$$EM_j = R^{-1}(F_j - X_k) \tag{4}$$

The Jacobian matrix of EM_j with respect to X_K is $H = [-R^{-1}\ 0_{3\times3}]$.

3) **Innovation**: the difference between measurement landmark and the observation is: $vc_{ij} = E_i - EM_j$, with covariance:

$$S_{ij} = HQ_{x_k}H^T + R^{-1}Q_{F_j}R + Q_{E_i} \tag{5}$$

where Q_{F_j} is the position covariance of the reference landmarks in the database in navigation frame, and Q_{E_i} is the covariance of currently reconstructed landmark in local vehicle frame.

4) **Mahalanobis distance**: the Mahalanobis distance between the reconstructed landmark E_i and reference landmark F_j is written as:

$$dist(E_i, F_j) = vc_{ij}^T S_{ij}^{-1} vc_{ij} < \lambda \tag{6}$$

If $dist(E_i, F_j)$ is less than a threshold $\lambda = \chi(0.05, 3)$ defined by χ^2 distribution table, landmark F_j is considered to be a possible correspondence of E_i. We might obtain a series of candidate correspondences for each landmark inside the confidence area.

5) **Joint compatibility**: when there are multiple landmarks being detected at the same time, instead of choosing the nearest neighbor of each landmark, the joint compatibility of all the road landmarks is taken into account. All the reconstructed landmarks E_i with at least one candidate correspondence are put into a single observation vector with uniqueness constraint, as:

$$\mathcal{E} = \{E_i\}^T = \{(\Delta X_i, \Delta Y_i, \Delta Z_i)\}^T, i \in 1, ..., m \tag{7}$$

The observations from different corresponding landmarks are also concatenated as:

$$EM = \{R^{-1}(F_{j_i} - X_k)\}^T \tag{8}$$

where F_{j_i} is the corresponding reference landmark of E_i. The Jacobian matrix of EM with respect to X_k is $H = -\left[R^{-1},\ ...,\ R^{-1}\right]_k^T$. Difference between the measurement and observation vectors is: $vc = \mathcal{E} - EM$. Covariance S of the vector vc is calculated the same as in Eq. 5. If the Mahalanobis distance $dist(\mathcal{E}, F)$ is less than threshold $\lambda = \chi(0.05, 3k)$, k being the number of landmark correspondences under uniqueness constraint, this combination of landmark association is considered as an acceptable correspondence. All the possible combinations of correspondences inside the gating area are kept and the matching ambiguities will be resolved by sequential matching. Then, the vehicle track is split into multiple tracks to maintain each landmarks association hypothesis with a parallel filter, as [15].

5) **Pose correction**: for each validated landmark association, the vehicle state in each track can be updated in parallel by Kalman gain:
$K = track(i).Q_{x_k} H^T S^{-1}$ and $\bar{X}_k = X_k + K \times vc$, and the pose uncertainty is updated to: $Q_{x_k} = (I - KH)Q_{x_k}$. For the vehicle positions without any visual landmarks in view, pose-graph optimization can be used to distribute the final pose correction to other vehicle positions without LOS (Line of Sight) of the visual landmarks in a local bundle adjustment in the future.

4 Evaluation

Experiments were conducted to test the proposed pose correction method with acquired data sequences. As presented in section 2, a ground Mobile Mapping System was used for data collection (Paris). The vehicle was equipped with a high-quality geo-referencing device (GPS/INS/odometer) and 12 rigidly installed

cameras on the roof of the vehicle, including a horizontal panoramic system of 8 cameras, 1 forward looking and 1 rear looking stereo pairs.

A data sequence of 2015 positions (about $12km$) was used to generate the landmarks database (orange trajectory in Fig. 5). As seen in Tab. 1 and Fig. 5, 120 road signs (yellow squares with red crosses) and 2116 road marking strips were generated with sub-decimeter accuracy. In average, at least one road sign exists for every 100 meters along the road. During the acquisition stage, the forward stereo pair can detect at least one road marking strip in front of the vehicle at about 50% locations. Road signs and road markings were respectively stored in a file of $351k$ and $890k$. The whole data volume per kilometer was $103k$.

Table 1. Statistic data of reference database

Images	Vehicle trajectory	Number of road signs	Number of marking strips
2015×12 *cameras*	12km	120 (351k)	2116 (890k)

We evaluated the contribution of road signs and road markings with another data sequence acquired by the same vehicle, but at different time. As the test path did not completely overlap with the reference sequence, we manually chose two portions of the test sequence which were long enough and the area of the path had been mapped in the previous stage (shown as cyan lines in Fig. 5).

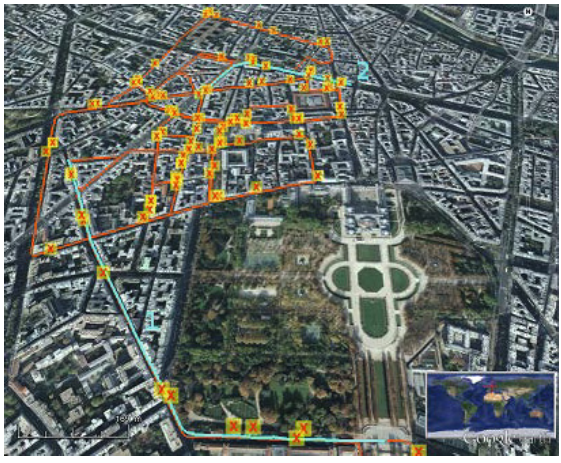


Fig. 5. Road sign landmarks (yellow squares) and vehicle trajectories overlapped on Google Earth. The reference data sequence used for generating the landmarks database is shown in orange; the two test trajectories are shown in cyan.

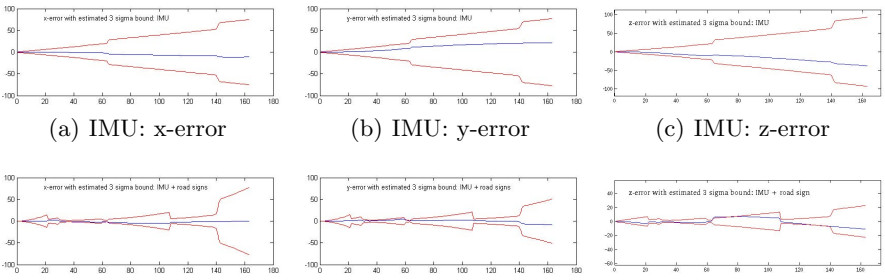
Due to an occasional cable connection problem of the front looking stereo pair in the map generation stage, no road markings were generated for the area of the

first segment (left bottom), thus this segments was with only road sign reference; the second segment was with both road signs and road markings. Pose ground-truth of the two segments were provided by GPS/INS/odometer post-processing software (though even this "ground-truth" might not be perfect, we will discuss this problem in the following experiments). Lengths of the two segments were respectively 1013m and 533m. The localization performance was evaluated using the number of true positive pose corrections, defined as:

- True positive (TP): landmarks were detected in images and associated with the corresponding database landmarks;
- False positive (FP): landmarks were detected in images, but associated with wrong database landmarks;
- True negative (TN): there was no corresponding landmark of a detection due to false detection or the incompleteness of the database;
- False negative (FN): landmarks were detected in images but not associated with the corresponding landmarks in database.

4.1 Localization Results of Segment 1

By assuming that the vehicle Euler angles were accurate and the vehicle initial position and velocity were known by GPS, the vehicle accelerations and rotations exported from the high-precision positioning system (with frequency of 100Hz) were used to predict the vehicle positions at first. Without any absolute measurements for pose correction, the vehicle trajectory drifts gradually, as seen in Fig. 6 (first row). Then, if a road sign is detected and associated, it is applied to adjust the vehicle trajectory, example is illustrated in Fig. 7; if there is no corresponding road landmark being detected for long period, the error ellipsoid of the vehicle position continues growing. Average linear distance between two



(a) IMU: x-error (b) IMU: y-error (c) IMU: z-error

(d) IMU+road sign: x-error (e) IMU+road sign: y-error (f) IMU+road sign: z-error

Fig. 6. Vehicle position error before (first row) and after (second row) incorporating road sign based correction. Blue curves: vehicle position error with respect to the ground truth; red curves: 3-sigma (3 times the standard deviation of the estimated position error)

detected road signs is 156m along the vehicle trajectory. As seen in Fig. 6 (second row), the position error after incorporating road sign based correction is in the form of sawtooth, and the average position error is reduced from 30m to 5.5m.

Fig. 7. Segment 1: Left: reference landmark overlapped on image frame with predicted vehicle pose; right: reference landmark overlapped on image frame after vehicle pose correction

Some statistic data of position correction with road landmarks is listed in Tab. 2. For segment 1, road signs were detected/reconstructed at 21 locations, 10 positions were adjusted by correctly associated road signs with the reference database, as shown in Fig. 8 (positions linked by red lines). Even with these limited information, the reference road landmarks still provide some useful corrections to the vehicle trajectories. 10 reconstruction were not associated with any landmarks and marked as true negative due to the incompleteness of reference database (3 reconstructions in this test as shown in Fig. 8 by yellow lines)

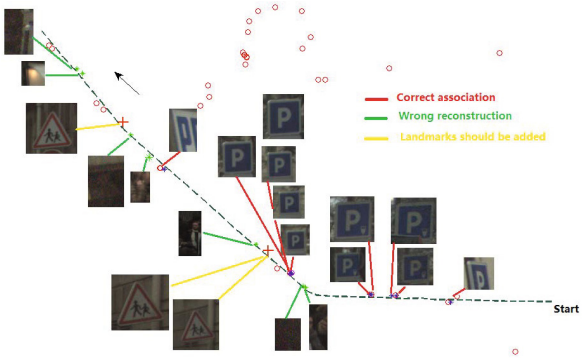


Fig. 8. Segment 1: landmarks association results (Red line: correct association; green line: wrong detection; yellow line: landmarks to be added into the database; red circles: reference road signs)

Table 2. Statistic data of position correction with road landmarks (Segment 1, with 165 vehicle positions in total; Segment 2: with 80 vehicle positions in total)

	Seg. 1	**Seg. 2**		
Landmarks	**Signs**	**Signs**	**Markings**	**Signs/Markings**
Locations with detections	21	15	50	59
TP (Correct association)	10	8	19	29
FP (Wrong association)	0	0	0	0
TN (No correspondence)	10	7	12	5
FN (Not associated with correspondence)	1	0	19	25

or wrong detection (7 reconstruction in this test as shown in Fig. 8 by green lines). But in reality, even though a road sign detection/recognition algorithm works perfect, as the road signs might be occluded by other vehicles along the street, it is difficult to obtain a complete landmark database by one acquisition.

4.2 Localization Results of Segment 2

For segment 2, we gradually added road signs and road markings for vehicle pose correction. When we displayed the onboard images of this segment using the poses provided by the GPS/INS/Odometer post-processing software, we observed that the images were consistent on horizontal dimensions, but did not overlap well in multiple runs on vertical dimension. Thus we only take use of the 2D positions as ground truth.

The vehicle positions predicted by accelerations were corrected by different landmarks. The position errors are compared in Fig. 9. As seen in Fig. 9, the average 2D position error of IMU based prediction was 6m, the error was 4.4m using IMU and road marking correction, the error was reduced to 2.46m using IMU and road sign correction. As road markings are more densely distributed in some area along the street, they can help to re-localize the vehicle more frequently, but also with more ambiguities especially on longitudinal direction. The

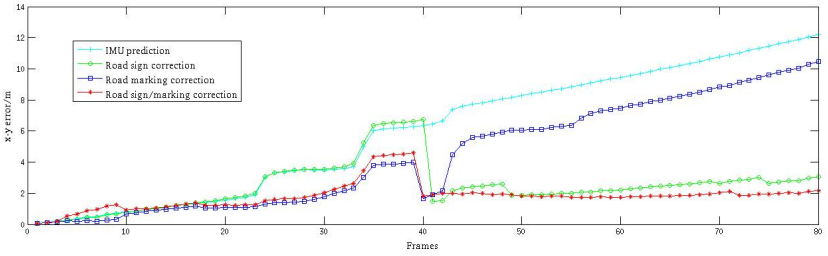
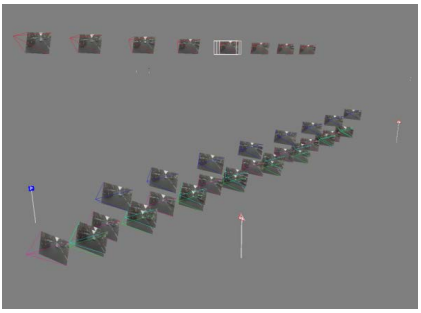


Fig. 9. Segment 2: Vehicle position error with IMU, IMU+road sign, IMU+road marking, IMU+road sign+road marking

distinctive road signs can help to improve the vehicle position precision on lateral and longitudinal directions. After incorporating road signs and road marking together, the error is further reduced to 1.81m. Although we don't have explicit ground truth of vehicle elevation, we noted some examples of elevation correction, like in Fig. 10, after incorporating the visual landmarks, image frames after vehicle pose correction are much more coherent with the database landmarks.



(a)

(b)

(c)

Fig. 10. Segment 2: a) From above to bottom: image frames predicted by IMU (in white box), GPS/IMU/Odometer software, IMU+road landmarks; b) Image frame predicted by GPS /INS /Odometer post-processing software; c) Corrected image frame after incorporating the road infrastructure objects

4.3 Complexity of the Method

The online localization processing is composed of pose prediction, landmarks (road signs/marking) detection/reconstruction, landmark association and vehicle pose correction steps. With current state-of-art techniques of landmark detection [16], it is possible to achieve real-time performance in the detection stage

(the techniques of road sign and marking detection we employ can be easily optimized to be real-time). Because the data volume of the infrastructure database is much more smaller than the popularly used point cloud (as in [4] for example), it is also possible to achieve real-time performance in the stage of landmark association: typically on the segments tested during this step, matching of road signs/markings involves the comparison of about 20 simple features at maximum, while point-based approaches would involve the manipulation of several hundreds of thousands of multidimensional features.

5 Conclusion

In this paper, we presented a road infrastructure database based vehicle pose correction method. Road signs and road markings were detected from forward-looking cameras and associated with the corresponding landmarks in the infrastructure database to correct the predicted vehicle pose. The experiments results demonstrated that the detected road signs/markings can be used as absolute control points to periodically adjust the vehicle positions. Although the proposed method aims to augment the vehicle localization performance in urban environments, it might also be applicable on rural roads. As the robustness of the whole system is affected by: 1) robustness of the road landmark detection/recognition algorithm; 2) since the same type road landmarks look exactly the same, the ambiguity problem might not be solved by only one road visual landmark when the pose uncertainty is too large. After long period of being lost (without any pose correction), other global localization methods should be adopted to re-initialize the vehicle global position, like place recognition method (with ten meters of accuracy as reported in [17]), vehicle trajectory and road network based absolute localization method, etc. Besides, vehicle positions without road signs in view might be adjusted by pose-graph optimization or a bundle adjustment. IMU can be replaced by camera based estimation as the research on visual odometry or structure-from-motion is more and more mature now.

References

1. Quddus, M.A., Ochieng, W.Y., Noland, R.B.: Current map-matching algorithms for transport applications: State-of-the art and future research directions. Transportation Research Part C: Emerging Technologies **15**(5), 312–328 (2007)
2. Sattler, T., Leibe, B., Kobbelt, L.: Fast image-based localization using direct 2d-to-3d matching. In: ICCV, pp. 667–674 (2011)
3. Lategahn, H., Schreiber, M., Ziegler, J., Stiller, C.: Urban localization with camera and inertial measurement unit. In: Intelligent Vehicles Symposium, pp. 719–724 (2013)
4. Royer, E., Lhuillier, M., Dhome, M., Lavest, J.: Monocular vision for mobile robot localization and autonomous navigation. International Journal of Computer Vision **74**(3), 237–260 (2007)
5. Soheilian, B., Paparoditis, N., Boldo, D.: 3d road marking reconstruction from street-level calibrated stereo pairs. ISPRS Journal of Photogrammetry and Remote Sensing **65**, 347–359 (2010)

6. Soheilian, B., Paparoditis, N., Vallet, B.: Detection and 3d reconstruction of traffic signs from multiple view color images. ISPRS Journal of Photogrammetry and Remote Sensing **77**, 1–20 (2013)
7. Schindler, A.: Vehicle self-localization with high-precision digital maps. In: IEEE Intelligent Vehicles Symposium (IV), pp. 141–146, June 2013
8. Schreiber, M., Knoppel, C., Franke, U.: Laneloc: Lane marking based localization using highly accurate maps. In: IEEE Intelligent Vehicles Symposium (IV), pp. 449–454, June 2013
9. Lategahn, H., Stiller, C.: Vision only localization. IEEE Transactions on Intelligent Transportation Systems **15**(3), 1246–1257 (2014)
10. Maldonado-Bascon, S., Lafuente-Arroyo, S., Siegmann, P., Gomez-Moreno, H., Acevedo-Rodriguez, F.: Traffic sign recognition system for inventory purposes. In: IEEE Intelligent Vehicles Symposium, pp. 590–595 (2008)
11. Segvic, S., Brkic, K., Kalafatic, Z., Stanisavljevic, V., Sevrovic, M., Budimir, D., Dadic, I.: A computer vision assisted geoinformation inventory for traffic infrastructure. In: 13th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp. 66–73 (2010)
12. Soheilian, B., Tournaire, O., Paparoditis, N., Vallet, B., Papelard, J.P.: Generation of an integrated 3d city model with visual landmarks for autonomous navigation in dense urban areas. In: IEEE Intelligent Vehicles Symposium (IV), pp. 304–309, June 2013
13. Olson, E., Leonard, J., Teller, S.: Fast iterative optimization of pose graphs with poor initial estimates. In: ICRA pp. 2262–2269 (2006)
14. Neira, J., Tardós, J.D.: Data association in stochastic mapping using the joint compatibility test. IEEE T. Robotics and Automation **17**(6), 890–897 (2001)
15. Nieto, J.I., Guivant, J.E., Nebot, E.M., Thrun, S.: Real time data association for fastslam. In: ICRA, pp. 412–418 (2003)
16. Houben, S., Stallkamp, J., Salmen, J., Schlipsing, M., Igel, C.: Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark. In: International Joint Conference on Neural Networks. Number 1288 (2013)
17. Zamir, Amir Roshan, Shah, Mubarak: Accurate image localization based on google maps street view. In: Daniilidis, Kostas, Maragos, Petros, Paragios, Nikos (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 255–268. Springer, Heidelberg (2010)