Integrated Metric-Topological Localization by Fusing Visual Odometry, Digital Map and Place Recognition

Shuai Yang[1(✉)], Rui Jiang[2], Han Wang[1], and Shuzhi Sam Ge[2]

[1] Nanyang Technological University, Singapore 639798, Singapore
{sayang,hw}@ntu.edu.sg
[2] National University of Singapore, Singapore 117576, Singapore
rui_jiang@u.nus.edu, samge@nus.edu.sg

Abstract. Visual odometry, map-assisted methods and place recognition are all popular approaches to localize a mobile vehicle from three different perspectives. Separate implementation of these methods may cause the localization system vulnerable due to the drift issue and local pose estimation of visual odometry, the on-road assumption and tough initialization of map-assisted methods and the discontinuous output of place recognition. In order to give full play to their advantages, an integrated localization strategy is presented in this paper, where metric data such as visual odometry measurement, a digital map and topological data of place recognition results are incorporated. Place recognition assists initialization process and provides topological place estimation at all times. Gaussian-Gaussian Distribution is used for visual odometry raw measurement representation such that the errors of odometry is appropriately modelled. By comparing similarities between the digital map and odometry trajectories, we then use map-assisted approach to correct odometry estimation. Finally, a mutual check gives a criterion for judging whether metric and topological results are sufficiently consistent. Experiment results show that the integrated system outperforms subsystems with mean localization error at 2.9 m on our self-collected dataset with off-road scenarios.

Keywords: Vision-based · Metric · Topological · Localization

1 Introduction

Over the past decades, vision-based approaches have been sufficiently explored in robotics and computer vision area due to camera's strength such as low cost, light-weight and data redundancy. Visual Odometry (VO) [10] as an alternative in mobile vehicle's localization, is the process of locally tracking the position of a vehicle using only consecutive images as the input. Due to the iterative principle, drift issue of VO in long range navigation is still yet to be solved. Specified to monocular systems, since the Monocular VO (MVO) is not able to recover the

A. Gruca et al. (eds.), *Man-Machine Interactions 5*,
Advances in Intelligent Systems and Computing 659, DOI 10.1007/978-3-319-67792-7_34

absolute depth of each feature, motion estimates and map structure can only be recovered up to scale [4].

In [3], we have presented a localization framework that uses an available digital map and the on-road assumption to reduce the estimation error of MVO or SVO. The framework has been designed based on Monte Carlo Localization, where VO is used to generate particles with specific probabilities, i.e. possible trajectories, and shape matching between trajectories and the digital map is used to further weight the particles. By assuming that the scale of MVO follows uniform distribution in the interval $[a, b]$, an Uniform-Gaussian distribution model has been proposed to handle the scale ambiguity in MVO. Although good localization results have been obtained from the presented approach on KITTI dataset, there are still some challenges as follows:

1. The map-assisted approach only works in on-road scenarios due to the on-road assumption. It is desired to make the approach applicable to both on-road and off-road situations since the vehicle does not always run on the road, or a dated map is used such that newly-built roads have not been added.
2. The vehicle's initial position and orientation are unknown and the initialization process relies extremely on shape matching performance. A very large number of initial particles needs to be generated to cover all possible trajectories (with different starting positions, orientations and scales). Thus, the initialization process may be time consuming. Occasionally, the initialization result converges to wrong locations due to similar road shapes.
3. After initial position estimation, it is more reasonable to model the scale distribution as Gaussian instead of uniform, as the optimal estimation of true scale must be a particular point instead of an interval.

Place recognition is the process of matching one query image against a database of geo-tagged images with known poses. It is actually an image retrieval task in computer vision field. Many localization approaches have been developed based on place recognition technique. The most well-known approaches are Fast Appearance-Based Mapping (FAB-MAP) [1,2,9] and SeqSLAM [6–8]. The former performs very large trajectory estimation based on a probabilistic framework, which is applicable even in visually repetitive environments. The latter uses sequential frame matching to find the best candidate and is applicable over extreme perceptual changes. Although the metric position of the query image can not be computed, a rough topological location can be obtained through place recognition operation. Noticed that a rough position information helps significantly to the initialization process of [3], it is promising to incorporate place recognition into our geometric map-assisted approach.

In this work, we aim at localizing a mobile vehicle equipped with one panoramic camera, one mono-camera and a digital map. Compared to the relevant work, the main contributions of the integrated approach are:

1. A sensor fusion strategy is proposed to combine metric data from digital map-assisted VO and topological data from place recognition results. Within the strategy, a mutual check thread is implemented to measure the accordance

between different data sources and to determine whether topological results and metric results should be trusted.
2. A robust place recognition aided initialization scheme is presented to initialize the localization framework and the initialization time consumption is significantly reduced.
3. Gaussian probability assumption instead of uniform assumption is used to represent scale distribution of MVO. The drift and scale ambiguity of SVO and MVO are modelled by Gaussian-Gaussian distribution more robustly.
4. An on-road/off-road judging scheme is proposed such that the integrated approach is applicable for both on-road and off-road scenarios.

2 Methodologies

In this section, topological and metric localization methods are explained separately first. Then an integrated framework is proposed based on their pros and cons.

2.1 Topological Localization Based on Place Recognition

Place recognition is one typical topological localization approach due to its discontinuous position estimation. Usually, there are two steps: database creation and online localization, involved in place recognition.

Database Creation. Lategahn et al. [5] introduced one visual feature, which they dub DIRD (Dird is an Illumination Robust Descriptor), among several millions that is best suited to represent places under illumination variations. Comparative experiments between DIRD and other descriptors demonstrated the effectiveness and efficiency of DIRD in place recognition. In this work, some changes are necessary to fit the particular application, and an improved DIRD is used to describe our panoramic images.

A vehicle equipped with one panoramic camera and a differential GPS (DGPS) travels the route to be recognized one or more times. As the vehicle travels the route, a database graph is created using the vehicle position at fixed distance intervals. Each node of the graph is annotated with the vehicle position and visual features. Vehicle positions are obtained from DGPS, while visual features are extracted from panoramic image. Both the vehicle position and visual features are stored in the database.

Thus, the database consists of the set $\mathcal{D} = \{f_k\}, k = 1, \cdots, K$ with components $f_k = \{\text{DIRD}_k, l_k\}$, where DIRD_k is the visual descriptor of the kth node; l_k is the ground truth location of the vehicle in the map.

Online Location Estimation. At run time, as the vehicle drives over the mapped routes, a search process is proposed to match the current panoramic

view I_t against the pre-stored database $\mathcal{D}$ through feature matching. A column vector $\mathbf{d}_t$ of L1 distances can be computed by

$$\mathbf{d}_t = (\|\mathrm{DIRD}_1 - \mathrm{DIRD}_t\|_1, \ldots, \|\mathrm{DIRD}_K - \mathrm{DIRD}_t\|_1)^T \tag{1}$$

Intuitively, the minimum argument of (1) can be considered as the result of the place recognizer.

However, matching the current query image with all the images stored in the database is quite time consuming. Furthermore, the place with the smallest L1 distance is not necessarily the best match due to dynamic objects, lateral shift or visual aliasing. To make our online localization scheme more efficient and reliable, several tricks are implemented. Firstly, a search window is used to restrict the matching range once place recognition has been initialized. Suppose f_{k_0} is the matching result at previous time step; w is the window size. This leads to a finite set $\mathcal{W}$ forming a sliding window around f_{k_0} placed in its center as:

$$\mathcal{W} = \{f_{k_0 - \frac{w}{2}}, \ldots, f_{k_0}, \ldots, f_{k_0 + \frac{w}{2}}\} \tag{2}$$

Then feature matching is only implemented inside this window and $\mathbf{d}_t$ becomes a vector of size w. The time consumption of this step is almost the same no matter how massive the database is. More importantly, the sequential consistency is maintained and positioning jump problem caused by perceptual aliasing is solved effectively. One thing should be noticed that global matching is still executed at intervals to correct failures of sliding window.

Secondly, a parametrized logistic function $\mathrm{logit}(\cdot)$ is used to convert all distances of $\mathbf{d_t}$ into matching probability $\mathbf{p_t}$ in the range $(0, 1)$. The logistic function is represented as

$$\mathrm{logit}(d) = 1 - \frac{1}{1 + \exp\left(-\alpha(d - d_0)\right)} \tag{3}$$

where d is one element of $\mathbf{d_t}$; d_0 is the L1 distance when the matching probability is 0.5 and α denotes the steepness of this sigmoid curve. Through this sigmoid function, large distance values are translated to small matching probabilities (near zero) and small distance values are translated to large matching probabilities (near one).

Lastly, it is noted that our panoramic camera has six small ones. Each camera has a constant field of view. It often happens that part of these cameras are filled with moving vehicles and pedestrians, which will cause partial appearance variations. If the panorama is considered as a whole, these partial appearance variations will severely disrupted feature matching performance. Instead, the images from each camera are considered separately in this work, so that partial appearance variations or aliasing will not determine the overall result. After computing the six probability vectors $\{\mathbf{p}_t{}^1, \ldots, \mathbf{p}_t{}^6\}$ from Eq. (3), their summation

$$\mathbf{p}_t = \sum_{n=1}^{6} \mathbf{p}_t{}^n \tag{4}$$

will be considered as the overall similarities. Then the node who has the highest matching similarity will be considered as the best matching result and the node's location

$$\mathrm{l}_k = \underset{f_k \in \mathcal{W}}{\operatorname{argmax}}\ \mathbf{p}_t \tag{5}$$

will be treated as the position estimation of place recognizer.

2.2 Metric Localization with Road-Constrained Visual Odometry Based on Gaussian-Gaussian Distribution

Compared with place recognition, the positioning result of visual odometry is a more accurate metric localization. Road-constrained VO is implemented to provide continuous pose estimation in [3]. In this work, most of the metric localization procedure follows our previous work [3] except that (1) Gaussian-Gaussian Distribution is used to generate possible MVO measurements; (2) Place recognition is implemented to assist initialization process.

Gaussian-Gaussian Distribution for Odometry Measurement Representation. Consider MVO measurement equation

$$\mathbf{t}_{k,k-1} = s_k \mathbf{t}^{\mathrm{raw}}_{k,k-1} \tag{6}$$

where $\mathbf{t}_{k,k-1}$ and $\mathbf{t}^{\mathrm{raw}}_{k,k-1}$ denote scaled translational vector and raw translational vector, respectively; $s_k \in \mathbb{R}^+$ is a scaling factor at time instant k.

In Monte-Carlo localization, improper model of MVO measurement may generate low-quality particles such that localization performance is degraded. Conventional methods usually regard $\mathbf{t}_{k,k-1}$ as Gaussian. In our previous work [3], we model $\mathbf{t}_{k,k-1}$ based on product distribution, where s_k and $\mathbf{t}^{\mathrm{raw}}_{k,k-1}$ are uniform-distributed and Gaussian-distributed random variables. Without a priori knowledge, it is reasonable to estimate scale s_k using uniform distribution. However, after obtaining the initial scale estimation, Gaussian-distributed s_k is preferred, as the true value of scale should be a point instead of an interval. In this paper, Gaussian-Gaussian distribution, which is used as MVO measurement model, can be defined as follows:

Definition 1 (Gaussian-Gaussian Distribution, GGD). *Given random variable S and random vector $\mathbf{X}$, the variate $\mathbf{Y} = S\mathbf{X}$ obeys Gaussian-Gaussian distribution $GG(ES, DS, E\mathbf{X}, \Sigma_X)$ if $S \sim N(ES, DS)$ and $\mathbf{X} \sim N(E\mathbf{X}, \Sigma_X)$, where $N(\cdot)$ denotes Gaussian distribution with corresponding expectation and covariance matrix (or variance).*

The expectation $E\mathbf{Y}$ and variance $D\mathbf{Y}$ of Gaussian-Gaussian distributed $\mathbf{Y}$ can be derived as

$$E\mathbf{Y} = ESE\mathbf{X} \tag{7}$$

$$DY_j = DX_j ES^2 + DSEX_j^2 + DSDX_j \tag{8}$$

where DS, DX_j and DY_j denote the variance of S, the variance of jth element in $\mathbf{X}$ and the variance of jth element in $\mathbf{Y}$, respectively.

Based on the above definition, the scaled MVO measurement $t_{k,k-1}$ can be represented with Gaussian-Gaussian Distribution $GG(ES, DS, E\mathbf{X}, \Sigma_X)$. Given the four parameters of GGD, samples denoting possible MVO measurements can be generated. With GGD, both scale ambiguity and measurement randomness are modelled simultaneously.

Similar to our previous work, a Monte-Carlo framework is leveraged to combine MVO measurement and geometric map. Each particle is considered as one possible trajectory of the vehicle. After map preparation and initialization, particles are generated from VO raw measurement. By comparing the trajectory of each particle with road shape obtained from geometric map through chamfer matching, weights are assigned to each trajectory. Resampling is implemented to retain trajectories with higher weights for position and scale estimation. The estimated scale will be used to generate particles in the next time step. Through the processes of this framework, visual odometry drift is corrected and scale ambiguity is eliminated.

Place Recognition Aided Initialization. Pure odometry based localization system could not possibly give a global position estimation without an accurate initial global position and orientation. An initialization scheme is proposed in [3], where a large number of initial particles is generated to cover all possible trajectories (with different starting positions, orientations and scales). And for each possible trajectory, one query edge map will be created. Then an exhaustive shape matching between the road edge map and all the query edge maps will be implemented to find the most possible localization hypotheses. Although experiments on KITTI benchmark had shown the effectiveness of shape matching based initialization, several challenges are still yet to be solved. First of all, shape matching performance depends on vehicle's trajectory and road conditions. Generally, the more complicate vehicle's trajectory is, the better performance will be. However, the vehicle's trajectory does not guarantee to meet such complexity. Pure road shape assisted initialization cannot ensure position convergence to true value. Moreover, the number of initial particles relates to the map size. The larger the map size is, the more initial particles are needed. Thus the initialization process becomes time consuming when the searching region is large.

In this work, the above challenges are solved by incorporating place recognition with road constrained approach. On the one hand, place recognition is firstly activated and the rough position estimation from place recognizer is then used to narrow down the initial searching region of the metric localization method. Only the nearby on-road area will be considered as the possible starting position. On the other hand, the road direction can be computed from the tangent orientation of the corresponding pixel point when generating road edge map. Assuming that initially the vehicle is parallel to the road, then the vehicle's starting orientation is always in accordance with the road direction. Hence, the number of the possible starting orientations is reduced dramatically. Initial particles are generated

to cover all the possible starting states. Then these particles will be fed into the shape matching scheme, from which the weight of each particle is obtained. A metric pose estimation will be computed after particle re-sampling. With this place recognition aided initialization, the large number of initial particles as well as the high requirement on road shape complexity in [3] are no longer needed. In other words, the initialization process becomes much easier.

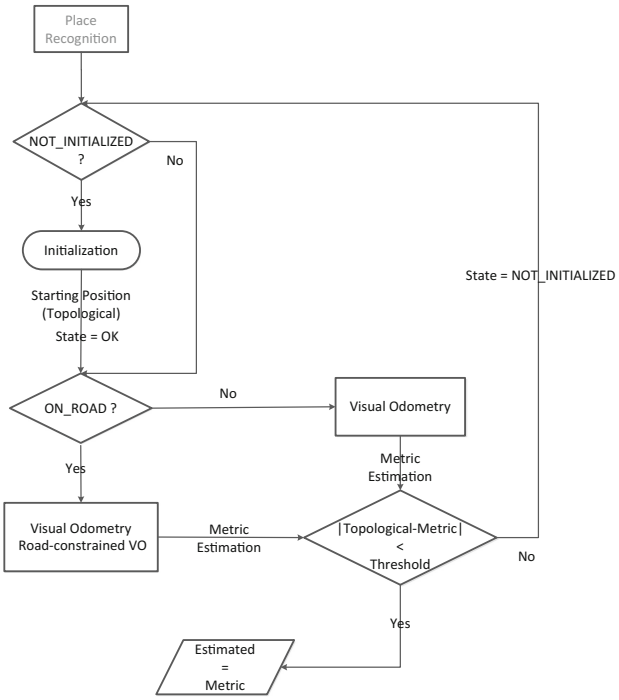

Fig. 1. The flowchart of the integrated vehicle localization method.

2.3 Integrated Localization Strategy

In previous sections, both topological and metric methods are described to localize a mobile vehicle. In order to give full play to their advantages, an integrated strategy is presented in this section.

Figure 1 demonstrates the flowchart of our integrated method, where no GPS or other sensors is involved, and only visual information and an OpenStreetMap is utilized to accomplish the localization goal. First of all, a state variable indicating the state of the whole framework is introduced. At each time step, this state is firstly checked. If it is "NOT INITIALIZED", place recognition aided initialization scheme explained in previous section will be performed. Once initialization is succeeded, the state variable will be assigned as "OK".

Another important feature in this framework is the on/off-road judging scheme, which makes the integrated approach applicable for both on-road and

off-road scenarios. When generating the panoramic database, on-road frames are labelled as one and off-road frames are labelled as zero. At run time, given one panoramic frame, the label of the best matched database frame will be considered as the current vehicle's state. If "ON ROAD" flag is false, only visual odometry is implemented since our road-constrained approach only works in on-road scenario. Otherwise, full road-constrained approach will be implemented. Afterwards, a mutual check thread is implemented to determine if the estimation difference between place recognition and road-constrained threads is smaller than a user defined threshold ϵ. If yes, the metric estimation will be considered as the final pose estimation. Otherwise, the initialization step will be re-executed.

3 Experimental Validation

3.1 Comparison Between GGD and UGD

Gaussian probability assumption instead of uniform assumption is used to represent scale distribution of MVO. This is the only improvement from GGD to UGD. In order to show their performance difference, evaluation experiments are conducted on KITTI benchmark. Scale estimation performance rather than the positioning result is leveraged to verify the advantage of the proposed localization frame. Some critical scale estimation results are shown in Fig. 2. The purple and red curves in the left are the upper and lower limits of the estimated scale interval from UGD. As can be seen, not all the estimated intervals cover the true scale. Since the scale estimation result is used for generating particles iteratively, once the scale interval $[a, b]$ estimated from the parameter estimation scheme does not contain the real one, the particle filter may diverge. The location estimation might be accurate but the scale ambiguity increases when the true scale is out of the estimated interval. This is an inherent flaw of UGD. To the contrary, GGD does not have this issue. The curves in the right show the estimated scale's expectation and variance from GGD. As can be seen, the expectation curves fit the ground truth well. The variance decreases rapidly at the beginning and becomes stable after a few iterations. Although, their are cases when the estimated means are a little far away from the true scales, particles generated in the next time step can still cover the truth with large probability as long as the true scales lie within three standard deviations of the mean. Thus, convergence can be ensured.

3.2 Localization Results

In order to evaluate the localization performance of the integrated strategy, experiments are conducted on our self-collected dataset.

Our evaluation mobile vehicle–Venus, is a self designed four wheeled mobile robot. It can be navigated by joystick at human walking speed. The robot is equipped with a DGPS to provide us with meter level position ground truth. A stereo camera set is mounted on the robot and oriented forward. It is configured

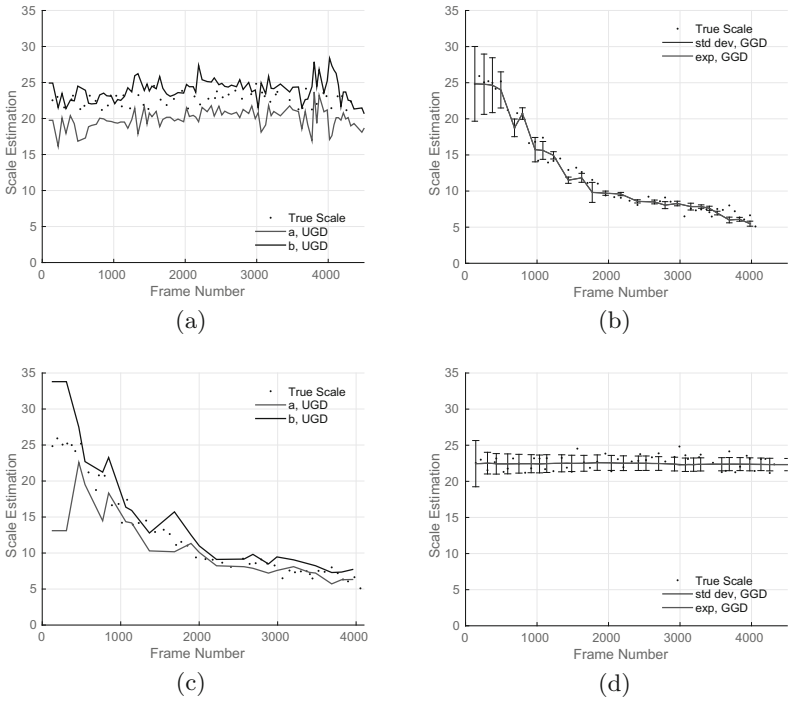

Fig. 2. Scale estimation comparison between UGD and GGD on KITTI 00 (up) and 08 (down). The left figures show the estimated scale's lower bound a and upper bound b of UGD. The right figures show the estimated scale's mean E_s and standard deviation σ_s of GGD. The ground truth scales are represented with black dots.

to acquire stereo frames at 10 Hz with a resolution of 1280×1024. The baseline of this stereo camera is configured at 30 cm to have a good effective depth range. The robot is also equipped with one Ladybug2 camera, which will be used to capture panoramic view images. Other sensors like IMU, laser range finder and compass are also available from this platform. All the sensors are configured by one CPU.

Our dataset was collected by driving Venus around the campus of Nanyang Technological University. The testing routes have two parts with a 3000 m length, including both on-road and off-road area. Figure 3 (left) is one screen shot of our experiment on route one. Trajectories estimated from DGPS, visual odometry, place recognition and the integrated approach are represented with different colors. As can be seen, visual odometry works fine in the first place, but it becomes worse and worse as the vehicle moves, especially when it comes to the off-road area. At the mean time, the positioning results from our integrated strategy are restricted to the road when the vehicle is travelling on the road. When the vehicle travels off the road, visual odometry takes over. A consistent good performance in both on-road and off-road areas is given from the proposed

integrated strategy. Noted that the positioning results from road-constrained approach are not plotted because they are the same with the results from the integrated approach when the vehicle travels on road.

Localization error curves from each of the above methods are demonstrated in the right side of Fig. 3. Place recognition updates at a low frequency while the other methods work continuously. The three circles show situations when re-initializations are activated due to the huge positioning difference between place recognition and road-constrained threads.

Table 1 lists the quantitative results of place recognition, visual odometry, road-constrained method and the proposed integrated approach, respectively. The whole position error of the integrated is less than 3 m over the 3 Km run, while the other methods either has a much larger error or has restrictions to use. Since all the roads in geometric map are modelled with centre lines, positioning error in the lateral direction of the road could not be eliminated. A 3 m positioning error is quite acceptable.

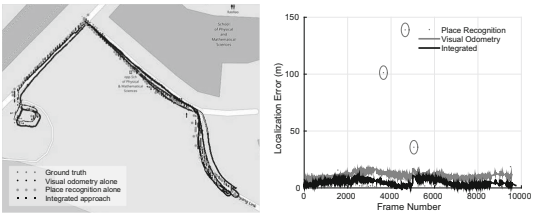

Fig. 3. Qualitative comparison and localization error.

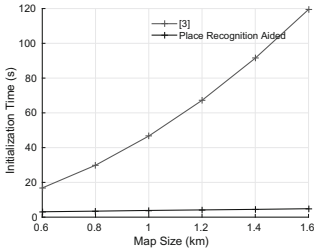

Fig. 4. Initialization time.

Table 1. Quantitative results of place recognition, visual odometry, road-constrained approach and our integrated strategy, respectively.

	Place recognition	Visual odometry	Road-constrained	Integrated
Avg error (m)	9.0	10.9	2.9 (only on road)	2.9
Median error (m)	2.7	10.7	4.6 (only on road)	4.3
Max error (m)	138.9	19.1	11.1 (only on road)	11.1

3.3 Initialization Analysis

In the particle filter framework, a lot of initial particles need to be generated to cover all the possible initial states. This makes initialization process be the most

time consuming part. In [3], the number of initial particles is largely determined by the map size. The red curve of Fig. 4 shows the relationship between the initialization time of [3] and the map size. It can be seen that as the map size increases, the initialization time grows in quadratic function. It is easy to understand this variation curve as the possible starting position increases in a square number when the map size increases. At the same time, the number of the initial states of the proposed place recognition aided initialization no longer depends on the map size. The purple curve of Fig. 4 shows the corresponding time variation. As can be seen, the proposed has a slightly increased initialization time due to the growing place recognition database. But the time consumption (around 3 s) is far less than [3]. The integrated approach is implemented in C++. And all the experiments run on a mobile workstation with an i7-4710MQ processor.

4 Conclusions

An integrated strategy has been proposed to localize a mobile vehicle equipped with one panoramic camera, one mono-camera and one digital map in this work. Place recognition, visual odometry and road-constrained approaches have been incorporated into one framework. With in this framework, place recognition plays a role of topological localization and assists initialization process. Road-constrained is responsible for on-road localization while visual odometry handles the off-road scenario. Gaussian assumption instead of uniform assumption has been proposed to model the scale distribution of monocular visual odometry. Evaluation results show that the proposed framework is highly accurate.

References

1. Cummins, M., Newman, P.: FAB-MAP: probabilistic localization and mapping in the space of appearance. Int. J. Robot. Res. **27**(6), 647–665 (2008)
2. Cummins, M., Newman, P.: Highly scalable appearance-only SLAM - FAB-MAP 2.0. In: RSS 2009, Sage Publications, Inc., Seattle, USA (2009)
3. Jiang, R., Yang, S., Ge, S.S., Wang, H., Lee, T.H.: Geometric map-assisted localization for mobile robots based on uniform-Gaussian distribution. IEEE Robot. Autom. Lett. **PP**(99), 1–1 (2017)
4. Kitt, B.M., Rehder, J., Chambers, A.D., Schonbein, M., Lategahn, H., Singh, S.: Monocular visual odometry using a planar road model to solve scale ambiguity. In: Carnegie Mellon University Research Showcase (2011)
5. Lategahn, H., Beck, J., Kitt, B., Stiller, C.: How to learn an illumination robust image feature for place recognition. In: IVS 2013. IEEE, Gold Coast, Australia (2013)
6. Maddern, W.P., Milford, M., Wyeth, G.: CAT-SLAM: probabilistic localisation and mapping using a continuous appearance-based trajectory. Int. J. Robot. Res. **31**(4), 429–451 (2012)
7. Milford, M.: Vision-based place recognition: how low can you go? Int. J. Robot. Res. **32**(7), 766–789 (2013)

8. Milford, M., Wyeth, G.F.: SeqSLAM: visual route-based navigation for sunny summer days and stormy winter nights. In: ICRA 2012, St. Paul, Minnesota, USA, pp. 1643–1649. IEEE (2012)
9. Paul, R., Newman, P.: FAB-MAP 3D: topological mapping with spatial and visual appearance. In: ICRA 2010, pp. 2649–2656, Anchorage, AK, USA. IEEE (2010)
10. Scaramuzza, D., Fraundorfer, F.: Visual odometry [tutorial]. IEEE Robot. Autom. Mag. **18**(4), 80–92 (2011)