Epipolar geometry-based ego-localization using an in-vehicle monocular camera

Haruya Kyutoku[1], Yasutomo Kawanishi[1], Daisuke Deguchi[1], Ichiro Ide[1], Hiroshi Murase[1]

[1]: Nagoya University, Japan
E-mail: kyutoku@murase.is.i.nagoya-u.ac.jp

Abstract Nowadays, development of driving support systems and autonomous driving systems have become active. Vehicle ego-localization using in-vehicle sensors is one of the most important technologies for these systems. Accordingly, various attempts to localize own vehicle from in-vehicle sensors have been made. In general, the estimation accuracy of the traveling direction is lower than in the lateral direction. Therefore, we propose a highly accurate method for ego-localization of the traveling direction based on epipolar geometry using an in-vehicle monocular camera. The proposed method makes correspondences between in-vehicle camera images and database images with location information, and calculates the location using locations annotated to the corresponding database images. However, there are many gaps due to the difference of speed and trajectory of vehicles even if the images are obtained along the same road. To overcome this problem, the distance between the input image and the database image is calculated by the distance metric based on the epipolar geometry and the local feature method. An experiment was conducted using actual images with correct locations, and we confirmed the effectiveness of the proposed method from its results.

Keywords Localization, Computer vision, Epipolar geometry, Monocular camera

1. INTRODUCTION

Accurate ego-localization of vehicles is essential for fully autonomous driving systems of the near future. Sophisticated navigation requires reliable and precise location information. Furthermore, accurate location is important for sharing of road information. For example, road closure information using probe data was provided following the Great East Japan Earthquake [1] and several great disasters[1].

For ego-localization, Global Positioning Systems (GPS) are used commonly in navigation. However, consumer GPS cannot provide accurate locations. In particular, errors become very large in urban areas where there are many tall buildings. There is improved Real-Time Kinetic (RTK) GPS which provides accurate locations, but it is not common as a consumer product because it is very expensive. Moreover, it is still ineffective when satellite signals are blocked (for example inside tunnels).

For accurate localization, many techniques using various sensors and databases have been researched [2, 3, 4, 5, 6]. They use databases constructed from images, 3D maps, and so on. However, the use of a front-facing monocular camera and GPS data as inputs is a unique and challenging task. Lane localization of the vehicle can be achieved relatively easily using well established line detection and tracking techniques. However, longitudinal localization along the direction of travel is a difficult and important problem. Moreover, even with highly accurate localization using several expensive sensors, the error in the longitudinal direction is large [7].

Thus, we have previously researched ego-localization along the direction of travel by calculating correspondences between input images and images in a city image database obtained from a front-facing monocular camera [8]. However, this research aimed to detect obstacles, and was used to acquire corresponding frames for subtraction. Therefore, in this paper, we accurately evaluate the ego-localization accuracy of frame correspondence, and propose the improvement of accuracy.

This paper is organized as follows: Section 2 describes our proposed method using image distance and a localization method with greater accuracy than the image spacing of the database. Experimental results are presented and discussed in Section 3. Section 4 summarizes this paper.

2. EGO-LOCALIZATION USING INPUT CAMERA IMAGE AND DATABASE IMAGES

The proposed method considers the image distance measure between input and database images. Section 2.1 describes the image distance measure in detail. The closest image can be obtained using the distance measure, but the accuracy depends on the image spacing of the database. Therefore, we present an improved ego-localization method with better accuracy than the database image spacing in Section 2.2.

2.1. Definition of the distance measure between two images

We define a distance measure for the positional relationship between two cameras using epipolar geometry [8]. When the

[1]Google,
https://www.google.org/crisisresponse/japan

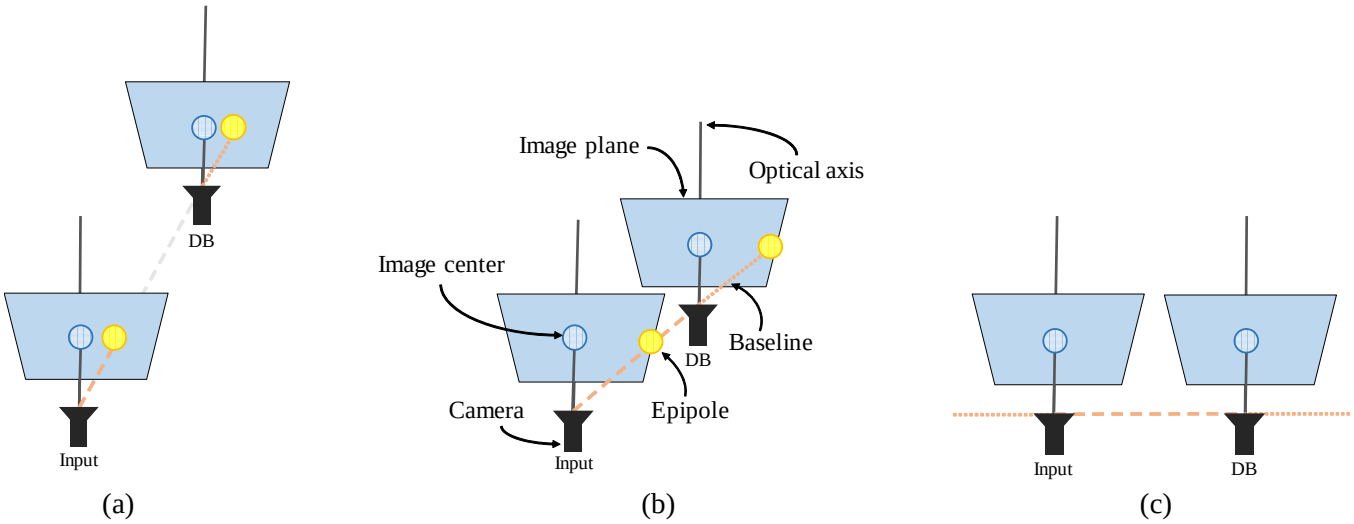

Fig. 1. Relationship between the positions of two cameras and their epipoles.

optical axes of the two cameras are almost parallel, the positions of their epipoles are strongly related to the distance between the two cameras, as shown in Fig. 1. Here, the epipole represents the position of the optical center of the second camera in the coordinate frame of the first camera. From Fig. 1, the position of the epipole moves away from the image center when the two cameras become closer. Figure 2 shows the positions of epipoles in actual images. An input frame is shown in the left column, and its closest frame $I_{i'}$ in the database is shown in the center of the right column. We can see that the position of the epipole changes according to the distance between the cameras.

The detail of the relationship between the cameras and epipoles is shown in a Fig. 3. Here, $\boldsymbol{O}$ is the optical center of the input camera, $\boldsymbol{Q}$ is the optical center of another (database) camera, f is the focal length of the input camera, and $\boldsymbol{E}$ is the position of the epipole. The horizontal and vertical pixel coordinates of the image plane are denoted x and y respectively, and z is the position along the optical axis, which is the traveling direction of the vehicle. From the figure, the relationship between the distance of the image center to the epipole e_x and the distance of the two cameras in the traveling direction $|q_z|$ is

$$|e_x| \quad = \quad \frac{f|q_x|}{|q_z|}. \tag{1}$$

Thus, $|e_x|$ is inversely proportional to $|q_z|$. In other words, the inverse of the distance of the image center and the epipole $1/|e_x|$ is proportional to the distance of two cameras with the traveling direction $|q_z|$, as

$$\frac{1}{|e_x|} \quad = \quad \frac{|q_z|}{f|q_x|}. \tag{2}$$

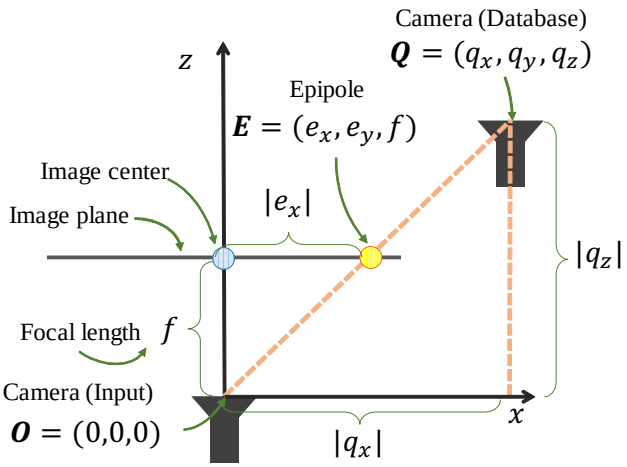

Fig. 3. Relationship between the position of the epipole and the distance between the optical axes of cameras.

Therefore, $1/|e_x|$ is defined as the distance metric between the images. It is possible to find the closest image in the database by using this image distance. Note that the accuracy of localization depends on the spatial interval of the database images.

2.2. Improvement of the localization using image distance

As described in Section 2.1, the image distance $d = 1/|e_x|$ is proportional to the distance of two cameras with the traveling direction $|q_z|$. So the image distance d satisfies

$$d \quad = \quad \alpha\,|q_z - l| \tag{3}$$

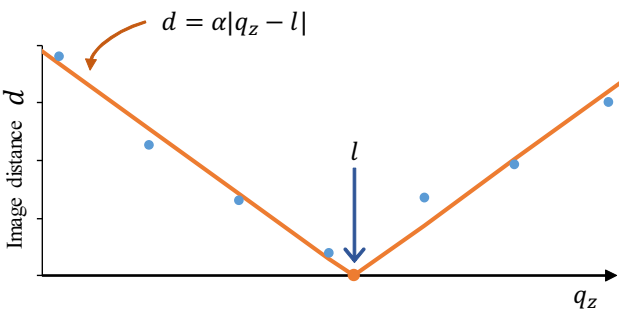

Fig. 4. Fitting using linearity of the image distance measure.

Fig. 5. Experimental vehicle.

where l is the correct position, and α is an arbitrary slope. The position l is calculated from the linear fitting of Eq. (3). The location is calculated by linear interpolation from location information corresponding to q_z before and after the position l.

3. EXPERIMENTS

To confirm the effectiveness of the proposed method, experiments were conducted by using actual images captured from an in-vehicle monocular camera.

3.1. Experimental conditions

To acquire in-vehicle images, we prepared a vehicle that mounted a monocular camera at the windshield as shown in Fig. 5. The used camera was Point Grey Research's Grasshopper3 GS3-U3-28S4C-C with Space VP-JHF8M-3MP, and the image size was $1,920 \times 1,440$ pixels. The vehicle traveled through the same 315 meters road in an urban area three times during the daytime. The route is shown in Fig. 6, and an example of an obtained frame is shown in Fig. 7.

Fig. 6. Route of experimental data.

Fig. 7. Example frame of the experimental data.

The correct positions of each frame were given by NDT matching [7]. To use NDT matching, a Velodyne's LiDAR HDL-64e was mounted on the roof as shown in Fig. 5, and obtained point clouds at the same time. Moreover, the positions obtained by the above processing were converted to the position in the traveling direction along the road, and were used as the correct position of each frame. This conversion was done with the Google Maps Roads API[2]. The positions of the epipole were calculated from a fundamental-matrix obtained by corresponding points with the RANSAC algorithm [9]. The corresponding points between input and database frames were obtained by using the SIFT feature [10]. Here, the frame rate of LiDAR is lower than that of the camera. So only images associated with point clouds were used for experiments, and as a result, the three sequences were composed of 579, 578, and 624 frames.

[2] https://developers.google.com/maps/documentation/roads/

3.2. Comparative method

In this experiment, the following methods were also tested for comparison.

1. GPS
   Output the mapped positions with the Google Maps Roads API from the corresponding GPS information. The used GPS was Javad's Delta 3, so results were more accurate than general GPS.

2. DTW (fixed)
   Output the positions of corresponding database frames. The corresponding frames are obtained by using the Dynamic Time Warping (DTW) method with the proposed image distance in Eq. (2). In this method, the start and end frame correspondences are fixed manually. Note that this method can not be applied for online processing since it needs fix the end frame.

3. Fitting (fixed)
   Output the positions by using the linear fitting described in Section 2.2. This fitting is performed around the result of DTW (fixed).

4. DTW (free)
   Output the positions of corresponding database frames just like DTW (fixed). In this method, only the start frame correspondence is fixed manually.

5. Fitting (free)
   Output the positions by using the linear fitting just like Fitting (fixed). This fitting is performed around the result of DTW (free).

Methods 2 to 5 were tested by applying each sequences as input or database.

3.3. Experimental Results

The result of each method is shown in Table 1. Minimum error, maximum error, mean absolute error, and standard deviation of error are displayed in meters. In addition, stacked charts of absolute error of each method are shown in Fig. 8. It shows the ratio of frames corresponding to each absolute error. In this graph, the upper and the more left the line is the smaller error is.

3.4. Discussion

From Table 1, all mean absolute errors using the distance metric based on the epipolar geometry were lower than that of an accurate GPS. Likewise, from Fig. 8, we can see that each method using the proposed distance metric is significantly more accurate than the GPS. Moreover, the results show that the fitting with Eq. (3) can reduce error. Additionally, we can see that the vehicle position could be estimated at 0.5 m or less with a probability of approximately 90 % by Method 5 capable of online processing. Therefore, the results suggest that the proposed distance metric works properly and effectively.

However, the maximum error of Method 5 is larger than that of Method 4. We consider that this was caused by the large maximum error of Method 4, because the fitting processing was performed around the result of DTW. Since the assumption of Eq. (3) is not satisfied at wrong positions, the fitting processing output wronger positions. Also, the calculation accuracy of the epipole position depends on the accuracy of the corresponding points. For this reason, it is necessary to examine the local feature with higher accuracy than SIFT.

Besides, the effect of the fitting was low in DTW with the fixed end point. We consider that this is because the correct positions made by the NDT matching originally included errors in the order of several tens of centimeters. Therefore, it is necessary to evaluate with more accurate position information.

4. CONCLUSION

In this paper, we proposed a highly accurate method for ego-localization of the traveling direction based on epipolar geometry using an in-vehicle monocular camera. In this framework, the position of the vehicle is calculated by image correspondence between an input and database images annotated with position information. The proposed method uses the epipolar geometry-based image distance metric. In addition, it estimates a more precise position by using the properties of this metric.

To demonstrate the effectiveness of the proposed method, an experiment was conducted by applying the proposed method to actual in-vehicle monocular camera images. Its results showed that the proposed method can estimate the positions accurately.

For future work, we need to consider a local feature with higher accuracy to calculate the accurate image distance. Furthermore, we should also evaluate with more accurate position information because the correct position used in the experiment contained some errors.

5. REFERENCES

[1] ITS Japan,
"Probe Helps Traffic Information in Disaster Area," `http://www.its-jp.org/english/its_asia/553/`, 2011.

[2] H. Uchiyama, D. Deguchi, T. Takahashi, I. Ide, and H. Murase, "Ego-localization using streetscape image

Table 1. Localization accuracy [m].

	1. GPS	**2. DTW (fixed)**	**3. Fitting (fixed)**	**4. DTW (free)**	**5. Fitting (free)**
Min	0.00116	0.00005	0.00006	0.00014	0.00031
Max	3.454	1.644	1.561	5.200	5.523
Mean	**0.691**	**0.193**	**0.185**	**0.296**	**0.250**
Standard Deviation	0.559	0.172	0.166	0.420	0.404

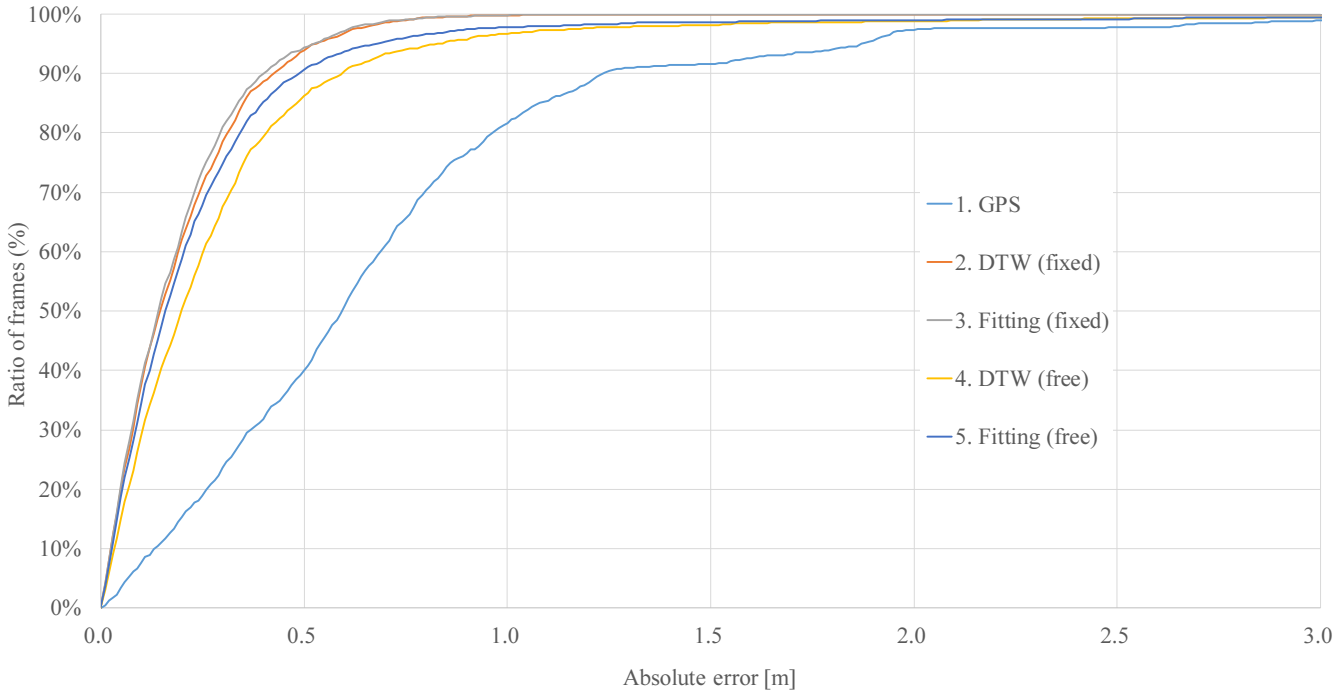

Fig. 8. Stacked chart of absolute error.

sequences from in-vehicle cameras," in Proc. *2009 IEEE Intelligent Vehicles Symposium*, pp. 185–190, 2009.

[3] M. Noda, D. Deguchi, T. Takahashi, I. Ide, H. Murase, Y. Kojima, and T. Naito, "Vehicle ego-localization by matching in-vehicle camera images to an aerial image," in Proc. *ACCV2010 Workshop on Computer Vision in Vehicle Technology: From Earth to Mars*, pp. 163–173, 2010.

[4] G. Ros, A. Sappa, D. Ponsa, and A. M. Lopez, "Visual SLAM for driverless cars: A brief survey," in Proc. *2012 IEEE Intelligent Vehicles Symposium Workshops*, 2012.

[5] R. W. Wolcott and R. M. Eustice, "Visual localization within LIDAR maps for automated urban driving," in Proc. *2014 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 176–183, 2014.

[6] R. W. Wolcott and R. M. Eustice, "Fast LIDAR localization using multiresolution Gaussian mixture maps," in Proc. *2015 IEEE Int. Conf. on Robotics and Automation*, pp. 2814–2821, 2015.

[7] N. Akai, L. Y. Morales, T. Yamaguchi, E. Takeuchi, Y. Yoshihara, H. Okuda, T. Suzuki, and Y. Ninomiya, "Autonomous driving based on accurate localization using multilayer LiDAR and dead reckoning," in Proc. *2017 IEEE Int. Conf. on Intelligent Transportation Systems*, pp. 1–6, 2017.

[8] H. Kyutoku, D. Deguchi, T. Takahashi, Y. Mekada, I. Ide, and H. Murase, "On-road obstacle detection by comparing present and past in-vehicle camera images," in Proc. *2011 IAPR Conf. on Machine Vision Applications*, pp. 357–360, 2011.

[9] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Comm. ACM*, vol.24, no.6, pp. 381–395, 1981.

[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol.60, no.2, pp. 91–110, 2004.

Line: Epipolar-Line, Intersection Point: Epipole.

Fig. 2. Position of epipoles between the input frame (left column) and its closest frame $I_{i'}$ and adjacent frames in the database (right column).