MVA2013 IAPR International Conference on Machine Vision Applications, May 20-23, 2013, Kyoto, JAPAN

Vehicle Localization along a Previously Driven Route Using an Image Database

Hideyuki Kume*
Nara Institute of Science and Technology
hideyuki-k@is.naist.jp

Arne Suppé, Takeo Kanade
Carnegie Mellon University
suppe@ri.cmu.edu, tk@cs.cmu.edu

Abstract

In most autonomous driving applications, such as parking and commuting, a vehicle follows a previously taken route, or almost the same route. In this paper, we propose a method to localize a vehicle along a previously driven route using images. The proposed method consists of two stages: offline creation of a database, and online localization. In the offline stage, a database is created from images that are captured when the vehicle drives a route for the first time. The database consists of images, 3D positions of feature points estimated by structure-from-motion, and a topological graph. In the online stage, the method first identifies the database image that is most similar to the current image by topometric localization, which considers topological information on a metric scale. The vehicle poses are then estimated from the 3D-2D correspondences of matching feature points between the current image and the identified database image. In an experiment, we estimated vehicle poses using images captured in an indoor parking lot.

1 Introduction

One of the fundamental requirements of an autonomous vehicle is the ability to determine its location on a map. Solutions to this localization problem frequently rely on GPS or expensive three-dimensional (3D) sensors [1]. Although GPS is a simple, low-cost solution, GPS signals are often unavailable due to occlusions in urban areas or indoors. 3D sensors are also used to help localize vehicles. However, 3D sensors are expensive, and may not be sufficiently durable due to rapidly moving internal parts.

In this study, in order to localize a vehicle, we use a camera that is inexpensive and can be used in many situations. We focus on the fact that in most autonomous driving applications such as parking and commuting, a vehicle follows a previously taken route, or almost the same route. Thus, we propose a method to localize a vehicle along a previously driven route by using an image database. The proposed method estimates the camera pose which is related to the vehicle pose by assuming that the transformation between the camera coordinate system and the vehicle coordinate system is known. As shown in Figure 1, the proposed method consists of two stages:

Offline creation of image database: An image database is created from images that are captured when the vehicle drives a route for the first time. This database consists of images, 3D positions of feature points estimated by structure-from-motion (SfM) [2], and a topological graph for topometric localization [3].

Online localization: The vehicle is localized using the image database and the current image by following three steps. First, we identify the database image that is most similar to the current image by topometric localization [3]. Next, we estimate the 2D-2D correspondences of feature points between the current image and the identified database image in order to obtain 3D-2D correspondences of the feature points for the current image. The vehicle pose is then estimated from these 3D-2D correspondences by solving the perspective-n-point (PnP) problem.

2 Related Work

In this study, we present only a small sampling of the approaches that are most related to estimating a vehicle pose along a previously driven route using an image database. These approaches can be classified into two types: topological localization-based methods and SfM-based methods.

2.1 Topological localization-based methods

Topological localization [3, 4] is a method that identifies the database image that is most similar to the current image. This method is efficient for vehicles that follow a previously taken route or almost the same route by considering topological information such as the spatio-temporal connections between sequential images. Some approaches combine topological localization and epipolar geometry [5] to estimate the current vehicle pose. Kosecka et al. [6] uses two-view geometry to estimate the relative pose between the current image and the identified database image. However, the scale of the relative pose cannot be estimated using two-view geometry. Murillo et al. [7] uses three-view geometry to estimate the relative poses between the current image and two database images (the identified image and its neighbor). In this case, the scale can be determined with respect to the distance between the camera positions of the database images.

The proposed method is also a topological localization-based method. Our contribution to existing topological localization-based methods is the use of SfM [2], which is more precise than three-view geometry.

2.2 SfM-based methods

SfM-based methods estimate the current camera pose from the 2D positions of feature points in the current image and their 3D positions obtained by SfM from database images. The problem of estimating camera pose from 3D-2D correspondences is well-known as

*This research was done while Hideyuki Kume was at the Robotics Institute of Carnegie Mellon University.

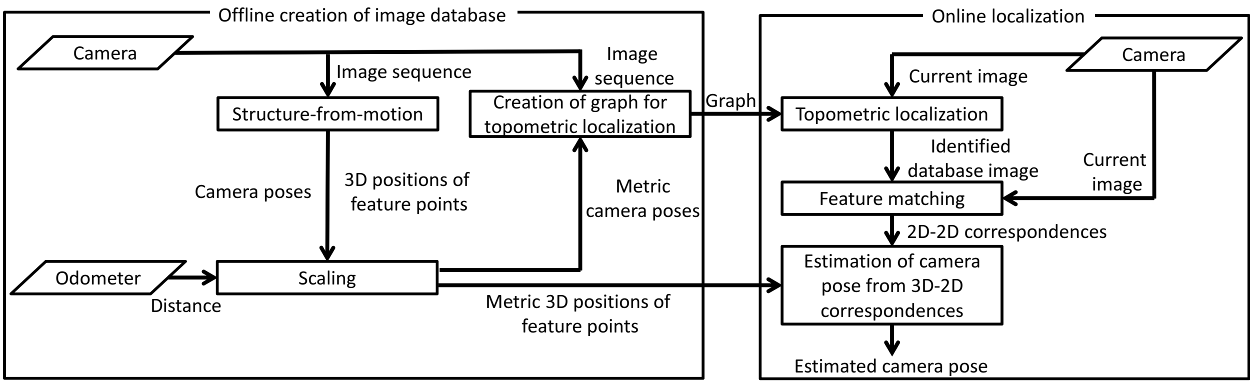


Figure 1. Flow of the proposed method

the PnP problem, for which there exist many solvers. The challenge that persists, though, is how to obtain the 3D-2D correspondences.

Taketomi et al. [8] estimates a tentative camera pose by tracking feature points temporally. Correspondences are then searched for only among those feature points that are projected onto the tentative camera's field of view. However, feature point tracking sometimes fails. Irschara et al. [9] uses image retrieval technique, which can identify the image that is most similar to the current image from among many images. Correspondences are then searched for only among those feature points that are in the identified image. However, image retrieval techniques only consider image features (e.g., SURF descriptors), and do not consider the topological information.

The proposed method is also an SfM-based method. Our contribution to existing SfM-based methods is the use of topometric localization [3], which is more efficient than the image retrieval technique by considering the topological information on a metric scale.

3 Offline Creation of Image Database

In our method, the database is created from images that are captured when the vehicle drives a route for the first time. The database consists of images, the 3D positions of feature points estimated by SfM, and a topological graph for topometric localization.

Structure-from-motion: We use VisualSFM [2], which is a state-of-the-art implementation of SfM, in order to obtain the 3D positions of feature points and camera poses. Since SfM cannot estimate the metric scale, which is required to estimate the vehicle pose on a metric scale in online localization, we measure the total driving distance using an odometer, which is standard equipment on ordinary vehicles. Specifically, the scale is determined by adjusting the total driving distance obtained by SfM based on the distance obtained by the odometer. In this way, metric 3D positions of the feature points and camera poses are obtained.

Creation of graph for topometric localization: We create a graph for topometric localization as in [3], except that the camera positions are estimated by SfM instead of GPS.

4 Online Localization

The vehicle is localized by using the image database and the current image in the following three steps.

Topometric localization: We first identify the database image that is most similar to the current image using topometric localization [3].

Feature matching: We next estimate the 2D-2D correspondences of feature points between the current image and the identified database image in order to obtain the 3D-2D correspondences of the feature points for the current image. We use SiftGPU [10], a GPU implementation of SIFT feature, to achieve real-time processing.

Estimation of camera pose from 3D-2D correspondences: In the feature matching stage, 2D-2D correspondences between the current image and the identified database image are obtained. We also estimated the 3D positions of the feature points for the database image in the offline stage. By combining these, we can obtain the 3D-2D correspondences of feature points for the current image. We then estimate the camera pose from these 3D-2D correspondences by solving the PnP problem. To do this, we simply use OpenCV's solver, which non-linearly minimizes the reprojection errors. We also apply RANSAC to reject outlier feature matches. If the number of inlier matches is smaller than a threshold, the estimated camera pose is ignored as a failure. In our experiment, the threshold was set to 6, which is the minimum number needed to solve the PnP problem linearly.

5 Experiment

In order to test the effectiveness of the proposed method, we evaluated its accuracy quantitatively by using image sequences captured in an indoor parking lot.

5.1 Experimental setup

We used a vehicle equipped with a sensor suite to evaluate our method. A camera was mounted on the roof of the vehicle. It was oriented approximately 45° to the right of a straightforward direction, and was configured to acquire 1024×768 pixel images. The vehicle was able to output its driving distance as with ordinary vehicles.

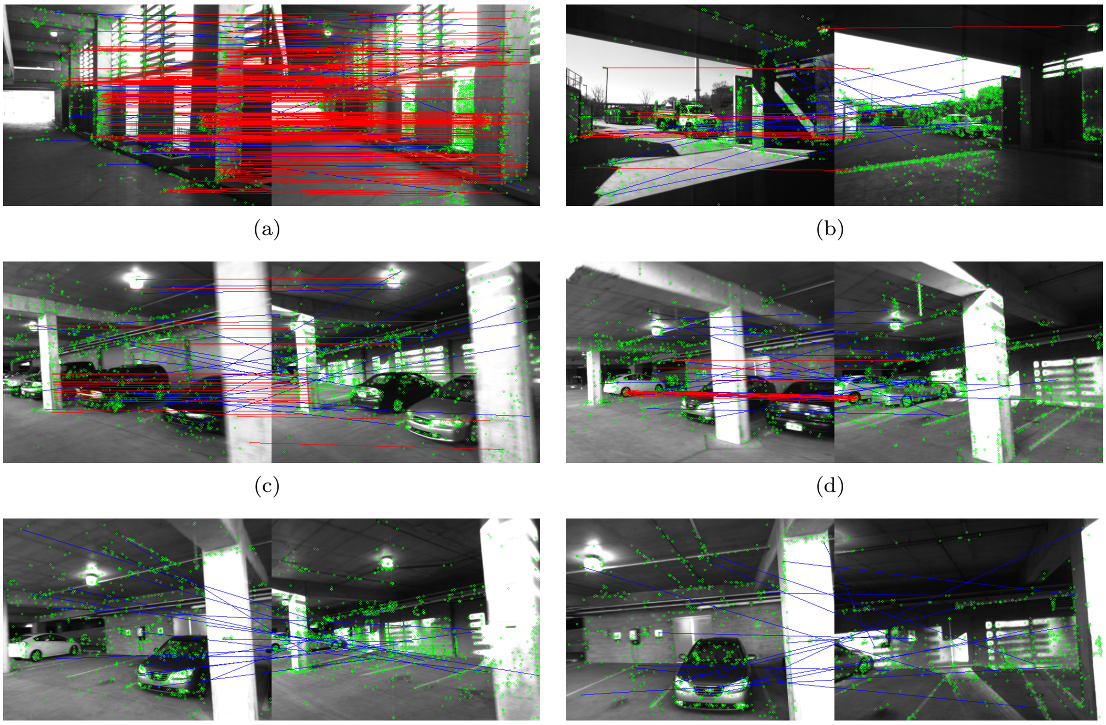


Figure 2. Examples of the input images (right), the database images identified by topometric localization (left), and the results of feature matching between these images (red line: inlier, blue line: outlier)

Two image sequences were captured in an indoor parking lot on different days by manually driving a route that describes a loop between an entrance and a parking space. One of the sequences was used for creating a database, and the other was used as input for online localization. Reference poses that were used to evaluate the accuracy of the estimated poses were estimated by offline SfM [2]. This offline SfM used feature matches among all the images and took a long time to achieve accurate estimation.

5.2 Localization results

Figure 2 shows examples of the input images, the database images identified by topometric localization, and the results of feature matching between these images. Figure 3 shows the vehicle poses and 3D positions of feature points for the database images, the estimated vehicle poses by the proposed method, and the reference vehicle poses. On the entire route, except for the last quarter, the estimated vehicle poses were almost the same as the reference vehicle poses despite changes in the illumination (Figure 2(b)) and the environment (Figure 2(c)). However, on the last quarter of the route, the estimated vehicle poses were unstable because the same car parked in a different place (Figure 2(d)). Failures also occurred due to errors in the topometric localization (Figure 2(e)) and significant changes in the illumination and the environment (Figure 2(f)).

Figure 4 shows histograms of the errors by comparing the estimated vehicle poses with the reference vehicle poses. Table 1 shows the average computation time of the proposed method, which was obtained using a PC with an Intel Core i7-2600k 3.40 GHz CPU and an NVIDIA GeForce GTX 580 GPU. From these results, it was confirmed that the proposed method estimated the vehicle pose at 8 Hz within a position error of 0.1 m and a posture error of 0.3° in approximately 70% of the input images. Although the proposed method sometimes provides unstable estimates and failures due to changes in the environment, it can be used in autonomous vehicles by combining those sensors that provide relative measurements such as odometry and IMU.

6 Conclusion

In this paper, we have proposed a method for localizing a vehicle along a previously driven route by using an image database that was created beforehand. The proposed method identifies the database image that is most similar to the current image by topometric localization, and it estimates vehicle poses from the 3D-2D correspondences of feature points between the database and the current image. Our experiment showed that the method estimated vehicle pose within a position error of 0.1 m and a posture error of 0.3° in approximately 70% of the input images captured in an indoor parking lot. In a future work, we will apply a Kalman filter in order to combine our results with odometry and IMU data for an autonomous vehicle application.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. EEC-0540865. Hideyuki Kume was supported by Young Researcher Overseas Visits Program for Vitalizing Brain Circulation of Japan Society for the Promotion of Science (JSPS).

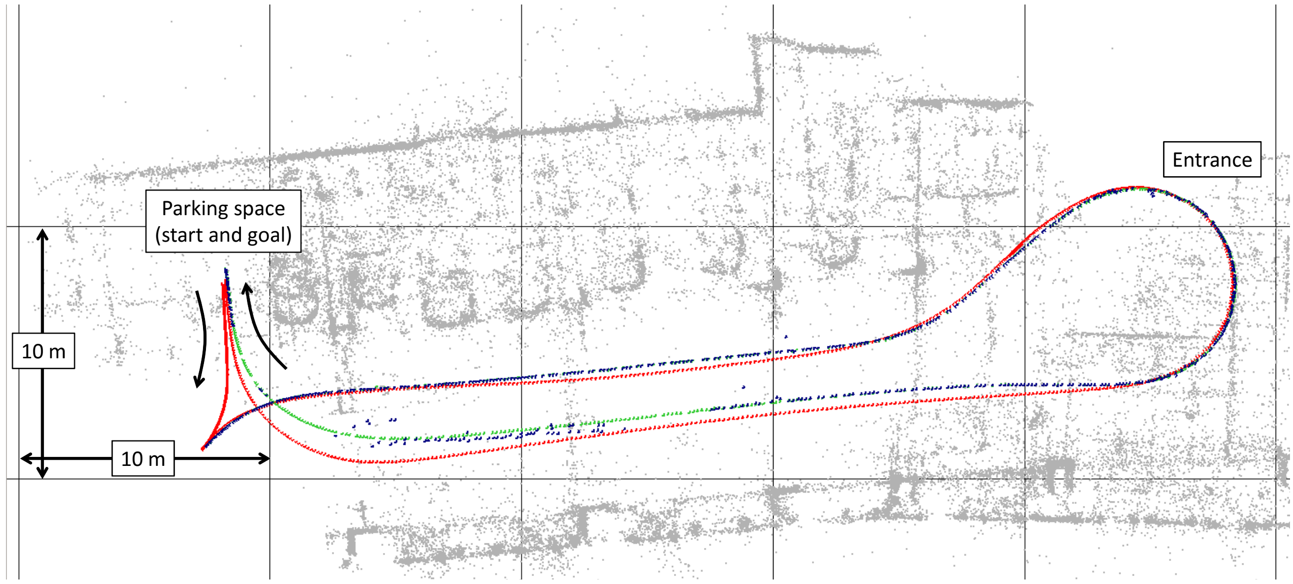
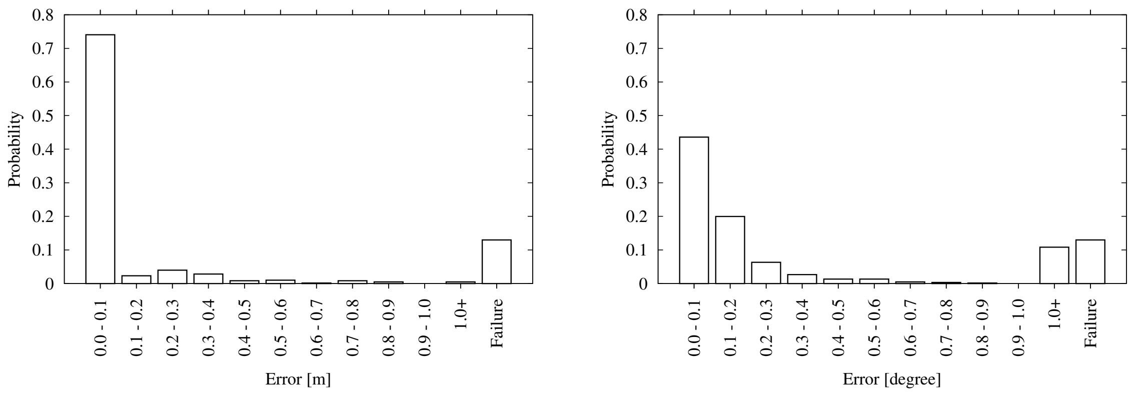


Figure 3. The vehicle poses (red) and 3D positions of feature points (gray) for the database images, the estimated vehicle poses by the proposed method (blue), and the reference vehicle poses (green)



(a) Position

(b) Posture

Figure 4. Histograms of the errors

Table 1. Computation time of the proposed method [msec]

Topometric localization	Feature matching	Solving PnP problem	Total
8.8	47.8	67.6	124.2

References

[1] S. Thrun et al., "Stanley: The robot that won the DARPA grand challenge," *J. of Field Robotics*, vol. 23, no. 9, pp. 661–692, 2006.

[2] C. Wu, "VisualSFM: A visual structure from motion system," http://www.cs.washington.edu/homes/ccwu/vsfm/, 2011.

[3] H. Badino, D. Huber, and T. Kanade, "Real-time topometric localization," in *Proc. Int. Conf. on Robotics and Automation*, 2012.

[4] C. Valgren and A. J. Lilienthal, "SIFT, SURF & seasons: Appearance-based long-term localization in outdoor environments," *Robotics and Autonomous Systems*, vol. 58, no. 2, pp. 149–156, 2010.

[5] R. I. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision," Cambridge University Press, 2004.

[6] J. Košecká, F. Li, and X. Yang, "Global localization and relative positioning based on scale-invariant keypoints," in *Robotics and Autonomous Systems*, 2005, pp. 27–38.

[7] A. Murillo, J. Guerrero, and C. Sagüés, "SURF features for efficient robot localization with omnidirectional images," in *Proc. Int. Conf. on Robotics and Automation*, 2007, pp. 3901–3907.

[8] T. Taketomi, T. Sato, and N. Yokoya, "Real-time and accurate extrinsic camera parameter estimation using feature landmark database for augmented reality," *Int. J. of Computers and Graphics*, vol. 35, no. 4, pp. 768–777, 2011.

[9] A. Irschara, C. Zach, J.-M. Frahm, and H. Bischof, "From structure-from-motion point clouds to fast location recognition," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2009, pp. 2599–2606.

[10] C. Wu, "SiftGPU: A GPU implementation of scale invaraint feature transform (SIFT)," http://cs.unc.edu/ ccwu/siftgpu, 2007.