Feature-Based 6-DoF Camera Localization Using Prior Point Cloud and Images

Hyongjin Kim[1], Donghwa Lee[1], Taekjun Oh[1], Sang Won Lee[1], Yungeun Choe[2], and Hyun Myung[1]

[1] Urban Robotics Lab, KAIST(Korea Advanced Insititute of Science and Technology), 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Korea
{hjkim86,leedonghwa,buljaga,lsw618,hmyung}@kaist.ac.kr

[2] Robot Research Lab, KAIST(Korea Advanced Insititute of Science and Technology), 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Korea
yungeun@kaist.ac.kr

Abstract. In this paper, we present a new localization algorithm to estimate the localization of a robot based on prior data. Over the past decade, the emergence of numerous ways to utilize various prior data has opened up possibilities for their applications in robotics technologies. However, challenges still remain in estimating a robot's 6-DoF position by simply analyzing the limited information provided by images from a robot. This paper describes a method of overcoming this technical hurdle by calculating the robot's 6-DoF location. It only utilizes a current 2D image and prior data, which consists of its corresponding 3D point cloud and images, to calculate the 6-DoF position. Furthermore, we employed the SURF algorithm to find the robot's position by using the image's features and the 3D projection method. Experiments were conducted by the loop of 510m trajectory, which is included the prior data. It is expected that our method can be applied to broad areas by using enormous data such as point clouds and street views in the near future.

Keywords: Localization, Feature Matching, Point Cloud, 3D to 2D projection.

1 Introduction

For the past few decades, the topic of autonomous mobile robot and UGVs(Unmanned Ground Vehicles) has been widely investigated [1, 2]. The most essential element of an autonomous robot navigation system is to identify the physical location of the robot, which in technical terms, is the localization problem.

Currently, there exist numerous methods proposed to solve the localization problem. However, there still remain several challenges. Although GPS(Global Position System) is generally used as a choice system to resolve the localization problem, there are some technical difficulties, mainly the occurrence of poor results under conditions of non-line-of-sight or in radio shadow areas of GPS. In order to overcome this complication, an integrated sensor that is a fusion of GPS and INS(Inertial Navigation System) is widely used [3]. However, the state of INS technology is imperfect as well, as there are difficulties in estimating global positions due to accumulative errors.

J.-H. Kim et al. (eds.), *Robot Intelligence Technology and Applications 2*,
Advances in Intelligent Systems and Computing 274,
DOI: 10.1007/978-3-319-05582-4_1,

Recently, as the use of prior information became remarkably easier through such programs as Google Street View [4], the global research endeavor to solve the localization problem using prior information has gained momentum. One notable study that has exploited prior images information employed an integrated sensor that was comprised of a fusion of GPS, INS, and positions of prior images [5]. Another research team reported the use of LADARs sensors to construct a prior point cloud, which constituted the foundation on which localization was conducted with LADARs sensors [6]. However, the high cost of LADAR served as a bottleneck for practical applications. Thus, another study dealt with the development of a system with a cheaper sensor, which utilized a pre-constituted prior point cloud, a low-cost camera, and odometry information to perform localization [7].

Therefore, this study proposes the use of existing prior data as a novel localization method that only necessitates a camera. The precedent study of this system is introduced [8] in the work-in-progress secsion. In this system, the features are extracted from prior images, and 3D coordinates are obtained from the prior point cloud. Similarly, features are extracted from the current image. Matching is performed with the features of the current image and those of prior images. Finally, a 3D projection algorithm is applied to estimate the robot's location.

2 Proposed Localization System

In this study, our approach utilizes only the current 2D image data and prior data. The processing steps of our proposed localization system are illustrated in figure 1. For preprocessing, the 2D features of prior images are extracted using the SURF(Speeded Up Robust Features) algorithm [9]. Furthermore, the 3D coordinates of the 2D features are also extracted using a ray tracing algorithm in the prior point cloud. After performing preprocessing, the 2D features of the current image are extracted on the fly. The features of the current image are matched to those of prior images using a RANSAC(RANdom SAmple Consensus) algorithm based on 2D homography. Then, the 3D coordinates of the matched features from preprocessing are projected to the 2D features of the current image to estimate candidate positions. Finally, the final position of the robot is estimated by median and mean filter among the candidate positions.

2.1 Preprocessing

In the preprocessing step, it is possible to reduce the operating time by performing SURF and ray tracing algorithm in advance using prior data. First, a survey vehicle that is equipped with costly devices such as LIDARs and camera collects prior data. The prior data consists of a point cloud and images. Each image contains 3D coordinates of a camera center in the point cloud at the time when the image was taken. Each image utilizes a SURF algorithm to extract its own feature descriptors. The 2D coordinates of the extracted features in the image are expressed as lines in the point cloud, using a pin hole camera model. The first meeting 3D point is found on the line formed by connecting the camera center, the feature point, and point cloud by utilizing a ray tracing algorithm. The concept of the ray tracing algorithm has been visually depicted in

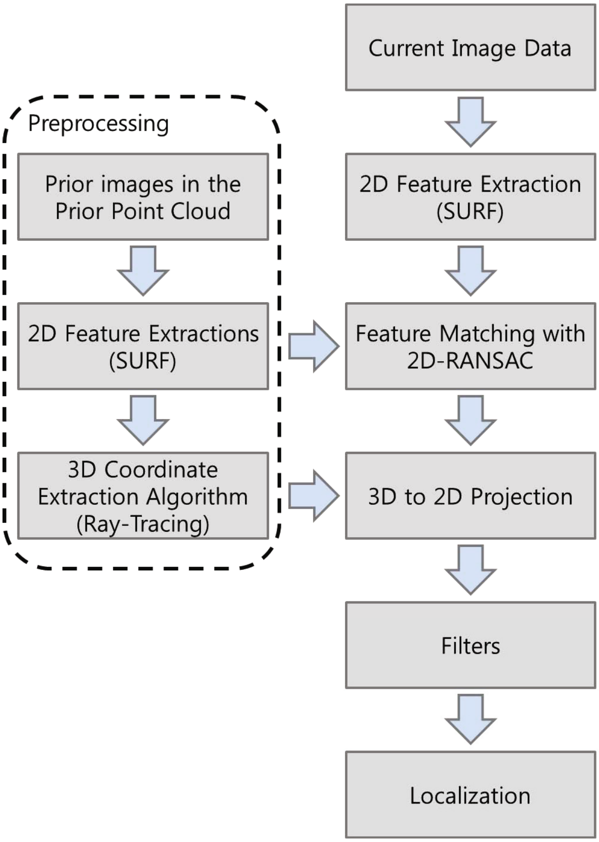
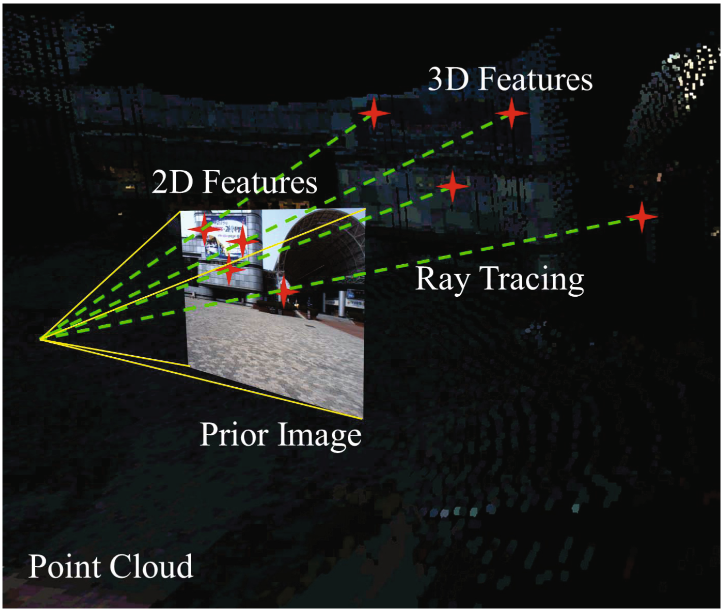


Fig. 1. Overview of the proposed localization system

figure 2(a), and figure 2(b) shows the ray tracing results for all images on the point cloud map. All feature descriptors of all prior images and its corresponding feature coordinates are saved for the main localization algorithm.

2.2 Feature Extraction, Matching, and 2D-RANSAC

In order to accomplish feature matching of the current image with prior images, the SURF algorithm is utilized. The intensities of brightness between the current image and prior images may show remarkable differences, even if the photos were taken from similar locations. This contrast can be attributed to the fact that the data were obtained from different cameras and time frames. Due to this discrepancy of light intensities, poor results are drawn in feature matching, even in similar places, as shown in figure 3(a). In order to solve this problem, all images udergo histogram equalization method before feature extraction. In figure 3(b), it is clear that the matching result is more successful than that without pretreatment. However, there is still much mismatching, due to the fact that the images were taken from different cameras and time frames.

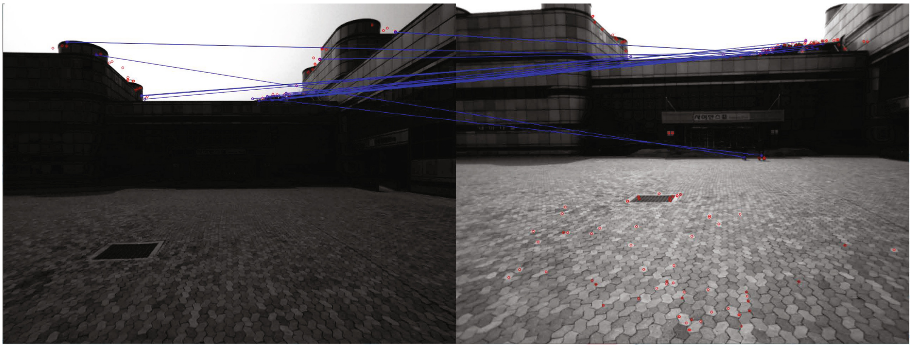


(a)

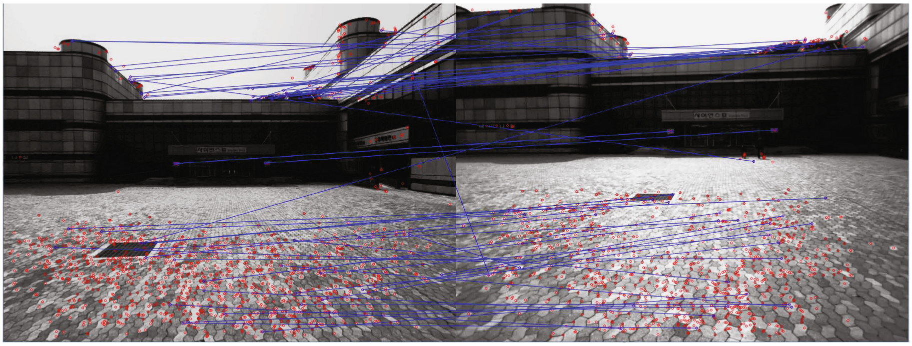
(b)

Fig. 2. (a) The green dashed lines show a concept of the ray tracing algorithm in the point cloud. The red stars indicate 2D and 3D features (b) The red lines indicate the ray tracing result of each 2D feature for a full path.

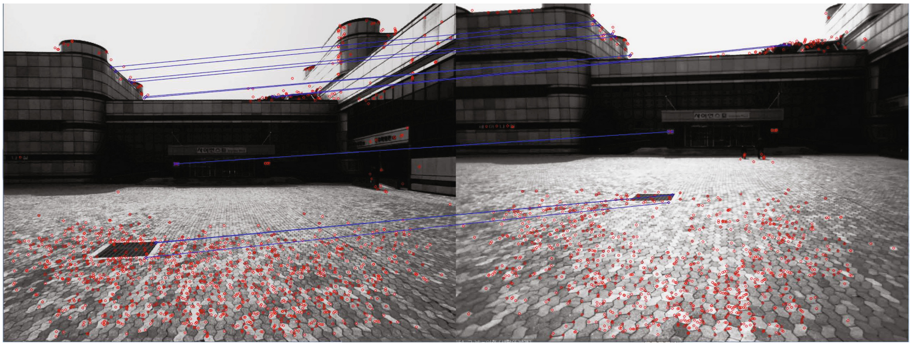
In order to eliminate the mismatching results, the RANSAC algorithm based on 2D homography was used. The final matching result is shown in figure 3(c). It is obvious that the outliers have been successfully removed. After carrying out feature matching of the current image and the prior images, the reference images are selected among the prior images by selecting the images that project points more than a predetermined threshold value.



(a)



(b)



(c)

Fig. 3. (a) Feature matching between a prior image(left) and a current image(right) without any pretreatment. (b) Feature matching after a histogram equlization method. (c) The result of the RANSAC algorithm based on 2D homography.

2.3 3D Projection Method

As the 3D projection method is essentially a mapping of 3D points to a 2D plane, the location of the camera center is estimated by mapping the 3D coordinates in the point cloud to a 2D image plane. The 3D coordinates in the point cloud are the values obtained by conducting the ray tracing algorithm of the feature points in the reference image. On the other hand, the 2D coordinates are the feature points that are matched with the current image. A 6-DoF(Degree of Freedom) position is estimated from a perspective projection method utilizing the parameters of a pin hole camera as follows:

$$[u, v, 1]' = K[R|T][x, y, z, 1]',$$

where K is the parameter of the pin hole camera model of the current image, R is 3-DoF rotation matrix and T is 3-DoF translation matrix. The 3D coordinates features in the point cloud from the reference image are expressed by homogeneous coordinates as $[x, y, z, 1]'$. The 2D coordinates features of the current images are also expressed by homogeneous coordinates as $[u, v, 1]'$. Between the reference image and the current image, one 6-DoF position is estimated. This is defined as the candidate position. As such, several candidate positions are estimated from the reference images that exhibit notable similarity to the current image.

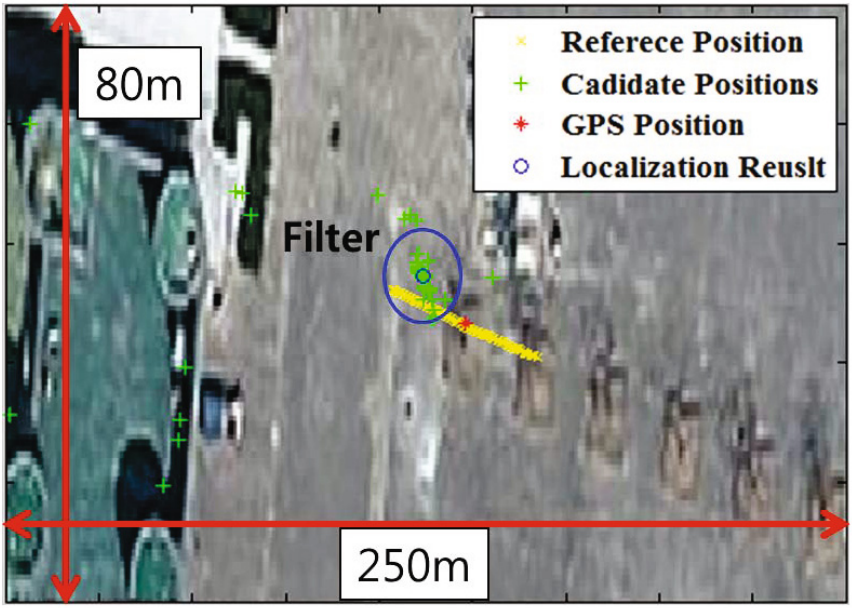


Fig. 4. The yellow x, green crosses, red star, and blue small circle indicate the positions of reference images, the candidate positions, the GPS positions as ground truth, and the final localziation result. The blue large circle indicates a boundary of filters.

2.4 Localization

There exist numerous false candidate positions which occur due to mismatching of features and ray tracing errors. To remove such incorrect results, median and mean filters are employed.

In this system, the median value of the geometric distances of candidate positions is calculated. This value is then used to determine outliers of the candidate positions by comparing a distance. If the number of inlier candidate positions is lower than a specific threshold, the localization is considered as a failure. Then, the final 6-DoF localization result is calculated by determining the mean value of inlier candidate positions. The result of a particular position is shown in figure 4.

3 Experiments

We have conducted experiments at the National Science Museum in Daejeon, South Korea. A survey vehicle was equipped with three SICK LMS-291 LIDARs, a camera, a Huace B20 GPS receiver, an Xsens MTi IMU, and a wheel odometry sensor. The camera had 1280 × 960 resolution and was mounted in the front. EKF(Extended Kalman Filter) sensor fusion of GPS, IMU, and odometry was used for localization in the survey phase. The prior data were obtained from one closed loop of 510m in the science museum. The prior data consisted of 1.0 million point cloud(x, y, z, r, g, b) and 3466 images. Each image had a 6-DoF position in the prior point cloud. The RMSE(Root Mean Squared Error) of the prior point cloud was lower than 1.5m.

To obtain the current data, a second vehicle was equipped with a Bumblebee XB3 stereo vision and a NovAtel OEM-Star GPS receiver. The current image was obtained from a left image of the stereo vision, which also has 1280 × 960 resolution. The GPS

Fig. 5. The red crosses and blue circles indicate GPS position and the result of localization for full path. The yellow lines show matched result between them.

Table 1. The result of accuracy

	X (m)	Y (m)
Standard deviation	3.22	2.18
RMSE	3.64	2.29

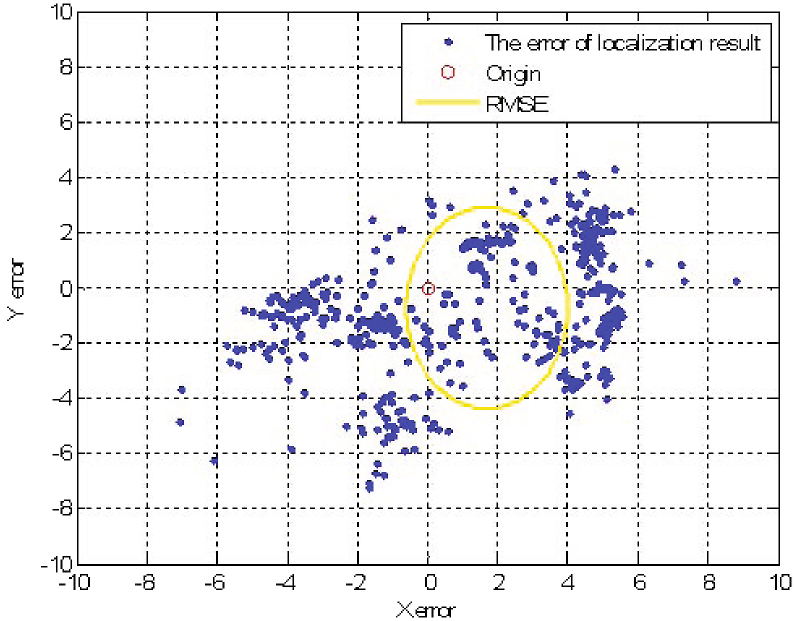


Fig. 6. The blue dots denote error of each localization result and the red circle denotes origin, and the yellow circle denotes RMSE of GPS data

data of 1.5m accuracy was used as a ground truth in this paper. The current data were obtained from the same closed loop used in the survey vehicle, by rotating around the loop three times. The result of 1,082 current images for all paths is shown in figure 5. The localization succeeded with a probability of 56%, a value which depended on the number of features on the images. The average Euclidean distance error of 606 successful localization points was 4.0m. The result of accuracy is shown in table 1. The error of each localization result is presented in the figure 6.

4 Conclusion and Future Works

In this paper, a novel localization method using a current image and prior data was proposed. The 3D features of the prior images were extracted in the preprocessing step. Reference images were searched for among the stack of prior images, by performing a correlative comparison of 2D features of the current image with the features of prior images. Furthermore, the candidate positions from reference images were obtained using a 3D projection method. Subsequently, the final positions were estimated by filtering of the candidate positions.

In this experiment, the RSME was comparably high than expected. This may be attributed to the erroneous 3D features due to inaccurate positions of the prior image in the point cloud. To resolve this problem, future studies will necessitate a closer investigation into the matching procedure between images and point clouds.

Acknowledgment. This work was supported by the R&D program of the Korea Ministry of Trade, Industry & Energy (MOTIE) and the Korea Evaluation Institute of Industrial Technology (KEIT). (The Development of Low-cost Autonomous Navigation Systems for a Robot Vehicle in Urban Environment, 10035354) The students are supported by Korea Ministry of Land, Infrastructure and Transport (MOLIT) as U-City Master and Doctor Course Grant Program.

References

1. Kanayama, Y., Kimura, Y., Miyazaki, F., Noguchi, T.: A stable tracking control method for an autonomous mobile robot. In: Proc. of IEEE Int. Conf. on Robotics and Automation, pp. 384–389 (1990)
2. Gage, D.W.: UGV history 101: A brief history of unmanned ground vehicle (UGV) development efforts. Unmanned Systems Magazine 13(3) (1995)
3. Obst, M., Bauer, S., Reisdorf, P., Wanielik, G.: Multipath detection with 3D digital maps for robust multi-constellation GNSS/INS vehicle localization in urban areas. In: Proc. of IEEE Intelligent Vehicles Symposium (IV), pp. 184–190 (2012)
4. Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., Weaver, J.: Google street view: Capturing the world at street level. Computer 43(6), 32–38 (2010)
5. Saito, T., Kuroda, Y.: Mobile robot localization using multiple observations based on place recognition and gps. In: Proc. of IEEE Int. Conf. on Robotics and Automation, pp. 1540–1545 (2013)
6. Baldwin, I., Newman, P.: Road vehicle localization with 2D push-broom lidar and 3D priors. In: Proc. of IEEE Int. Conf. on Robotics and Automation, pp. 2611–2617 (2012)
7. Stewart, A.D., Newman, P.: Laps-localisation using appearance of prior structure: 6-DOF monocular camera localisation using prior pointclouds. In: Proc. of IEEE Int. Conf. on Robotics and Automation, pp. 2625–2632 (2012)
8. Kim, H., Oh, T., Lee, D., Choe, Y., Chung, M.J., Myung, H.: Mobile Robot Localization by Matching 2D Image Features to 3D Point Cloud. In: Proc. of IEEE Int. Conf. on Ubiquitous Robots and Ambient Intelligence, pp. 266–267 (2013)
9. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)