https://doi.org/10.2352/ISSN.2470-1173.2018.09.AVM-283
© 2018, Society for Imaging Science and Technology

Vehicle Pose Estimation from Drive Recorder Images by Monocular SLAM and Matching with Rendered 3D Point Cloud of Surrounding Environment

Akiyoshi Kurobe, Hisashi Kinoshita, and Hideo Saito; Keio University; DENSO Corporation; Keio University

Abstract

Vehicle pose estimation is a vital technology for reconstructing the circumstaces of traffic accidents. We propose a novel method for reconstructing the trajectory of vehicles from drive recorder images and a point cloud around the road. First, we apply ORB-SLAM to image sequence of the drive recorder for obtaining the vehicle pose trajectory; however this is based on relative coordinates and a relative scale. For estimating the absolute coordinates and scale of the trajectory, which cannot be obtained from a monocular SLAM like ORB-SLAM, we match the feature points detected in the image sequence with the three-dimensional (3D) point cloud of surrounding environment.

For finding 3D points matching the feature points, we generate candidate images by the rendering 3D point cloud of the surrounding environment using the position initially estimated by the Global Positioning System (GPS). Next, we match to obtain the 3D two-dimensional (2D) generated images and drive recorder image to get 3D-2D point correspondences between the 3D point cloud and the drive recorder images; thus, we can convert the relative estimation of the camera pose by ORB-SLAM to the coordinates of the 3D point cloud of the surrounding environment. In the evaluation experiments, we confirmed the effectiveness of our method by comparing the vehicle poses estimated by our method, with those of RTKGPS, which exhibits high measurement precision.

Introduction

In recent years, along with aging, traffic accidents of elderly drivers and buses have occurred frequently, and there have been many victims of such incidents. Therefore, in recent years, driving support system [7, 9, 15], automatic braking systems for accident avoidance [17, 6, 4], and automatic driving technology [16, 19, 11] have been actively researched. For the practical application of these technologies, the trajectory drawn by the vehicle represents essential data.

The on-the-spot inspection of these accidents is currently limited to the method of reconstructing the vehicle's trajectory based on the brake marks left on the road and the damage to the surroundings. There are two problems with such approaches. One is that the accident cannot be reconstructed accurately, as the situation is affected by the roads and surrounding conditions before the accident. The other is that quantitative estimation results cannot be obtained.

We propose a novel method of estimating the vehicle's trajectory from the drive recorder videos and high-precision LiDAR point cloud to make practical use of these technologies for realizing more accurate on-the-spot inspection. In our proposed method, a drive recorder image sequence is input to the ORB-SLAM [13] to estimate the vehicle's trajectory. This trajectory is a relative estimation in the coordinate system based on the ORB-SLAM processing system [13]. However, in real on-the spot inspection and collection of trajectory data, it is essential to estimate the vehicle's trajectory on the real-world scale. That is, it is necessary to match the LiDAR point cloud with the drive recorder image. Thus, instead of directly calculating the relationship between the three-dimensional (3D) LiDAR point cloud and the two-dimensional (2D) drive recorder image, we generate candidate images by rendering the 3D point cloud of the surrounding environment using initially estimated position by GPS. Next, we match those generated images and drive recorder images to obtaind the 3D-2D point correspondences between the 3D point cloud and the drive recorder images, so that we can convert the relative estimation of the camera pose by ORB-SLAM [13] to the coordinates of the 3D point cloud of the surrounding environment.

In the evaluation experiment, we applied the proposed method to a drive recorder mounted on a vehicle traveling about 40 m. Regarding the position in the estimation results, we compare it with the highly accurate Real Time Kinematic GPS (RTKGPS), and for the posture, we generate the image from the point group based on the estimation results and compare it with the corresponding in-vehicle camera image to verify the accuracy.

Related Works

SLAM(Simultaneous Localization and Mapping) and VO(Visual Odometry) are essential technologies for autonomous driving. SLAM can be divided into RGB-D SLAM and Monocular SLAM. The latter monocular SLAM and VO do not require distance information and enable the camera trace and 3D reconstruction with only RGB images. However, these techniques have a disadvantage in that the scale can not be uniquely estimated from only RGB images. Thus, to solve scale uncertainty, Wolcott et al. [18] proposed a method of localizing an autonomous driving vehicle in urban environments. Using LiDAR intensity values, they render a synthetic view of the mapped ground plane and match it against the camera image by maximizing the normalized mutual information. Futhermore, Caselitz et al. [5] addressed the scale uncertainty by taking the optimization problem based on the ICP algorithm for point cloud reconstructed by ORB-SLAM [13] and LiDAR point cloud. Our method employs visual odometry to track the camera trajectory via local bundle adjustment. For this purpose, we rely on components of Stereo ORB-SLAM [13] presented by Mur-Artal et al.. This is a state-of-the-art, open-source solution for monocular SLAM that stands in a line of research with PTAM

[10].

Proposed Method

In this section, we describe our proposed method, which estimates vehicle's trajectory from drive recorder images and LiDAR point cloud data. Figure. 1 shows a flow diagram of the proposed method. Our method consists of three parts(ORB-SLAM, the matching rendered LiDAR point cloud and drive recorder images, and scale conversion). After 3D LiDAR, we describe each part in detail.

3D-LiDAR

Figure. 2 shows the 3D-LiDAR that we use to measure the 3D point cloud of the surrounding environment. This device can acquire color information, so we can obtain both depth and color information of each point. In actual measurement, this device is installed at intervals of several tens of meters on the road, and the point clouds are integrated in consideration of each measurement position.

ORB-SLAM

ORB SLAM [13] is a monocular SLAM using ORB feature [14] that can be computed at high speed by binary string description. the processing is divided into three threads, as follows: the tracking, matching and loop closure threads. At the beginning of processing, the scale is determined by initialization for the input RGB images. Next, based on the initialization scale and coordinate system, the relative camera pose is estimated, and the environmental map is reconstructed as a 3D point cloud. In our proposed method, it is necessary to estimate the relative vehicle's trajectory. For this purpose, we rely on the components of ORB-SLAM presented by Mur-Artal et al. [13]. The image taken by the drive recorder set in the car is affected by distortion of the lens of the drive recorder and the windshield. Thus, we calibrate the drive recorder and input drive recorder images to ORB-SLAM [13], with the distortion removed. Figure. 3 shows the drive recorder images before and after distortion removal. With this process, we can estimate the relative vehicle's trajectory.

Matching the Rendered LiDAR Point Cloud and Drive Recorder Images

Via the processing in the previous section, the relative vehicle's trajectory based on ORB-SLAM [13] is estimated. To convert the relative trajectory to the scale of the real scale LiDAR point cloud, we need match the point cloud and drive recorder images. However, it is difficult to directly perform this matching, so we match the image generated from the point cloud and the drive recorder image. It is assumed that the location where each frame of the drive recorder image is taken in the point cloud has already been acquired by the GPS mounted in the vehicle. However, since the general GPS includes an error of several meters, a more accurate estimation of the vehicle position and orientation is necessary. In this proposed method, we generate images from the LiDAR point cloud based on the position obtained from the GPS and the direction of the vehicle.

Generate Images from Drive Recorder Image

We generate projection matrix $\boldsymbol{P}$ in Eq. (1) from the information on the GPS position and the rough direction of the vehicle. By applying matrix $\boldsymbol{P}$ to the LiDAR point cloud, we convert each point cloud into a coordinate system with the virtual camera viewpoint:

$$\boldsymbol{P} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \boldsymbol{R} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{1}$$

$$\boldsymbol{R} = \begin{pmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\phi & \theta & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}. \tag{2}$$

As shown in Fig. 1, some of candidate images are generated by randomly changing the angle of Eq. (2) in accordance with the advancing direction. Next, a point group in front of the vehicle is projected onto the determined image plane. If a point cloud is projected to the same pixel, the point closer to the camera point of view is rendered. Figure. (10) shows an example of an image generated from the LiDAR point cloud.

Matching

Multiple viewpoint candidate images are generated from GPS information. In the next step, these generated images are matched with drive recorder images based on the RANSAC robust estimation method [8]. We calculate all the corresponding points and distance of corresponding. AKAZE feature [1] is used for matching between the generated candidate images and drive recorder images. The AKAZE feature [1] employs an algorithm that improves the KAZE feature's processing speed [2]. The Gaussian filter used in the SIFT [12] and SURF features [3] blurs the edges of objects, and it has the disadvantage that local features can not be extracted. To remedy this disadvantage, the KAZE feature [2] is processed so that local features can be extracted pace; thus, the calculation takes time. Thus, the AKAZE feature [1] uses a unique descriptor called a Modified-Local Difference Binary (M-LDB), and by incorporating unique ideas for speeding up the calculation of the pyramid structure, it is possible to improve the robustness and speed. We have learned empirically that we can even match robustly images that do not have many feature points robustly, like images generated from the point cloud.

Figure. (5) shows the matching between an image from which distortion has been removed and an image generated from a point group. With this matching, we can acquire a plurality of correspondences between the image coordinates and 3D point cloud. By solving the PnP problem from these corresponding points, we can estimate the position and orientation of the target drive recorder image in the LiDAR point cloud.

Transformation of the Coordinate System

The trajectory estimated by ORB-SLAM is a relative scale based on initialization. Therefore, it is necessary to convert the coordinate system of the ORB-SLAM to the coordinate system of the LiDAR point cloud. As shown in Eq. (3, 4), the first frame estimation result from ORB-SLAM is initialized (with the position and orientation of the camera) as the origin:

$$\boldsymbol{t}_{ORB\_init} = \boldsymbol{O}. \tag{3}$$

$$\boldsymbol{R}_{ORB\_init} = \boldsymbol{I}. \tag{4}$$

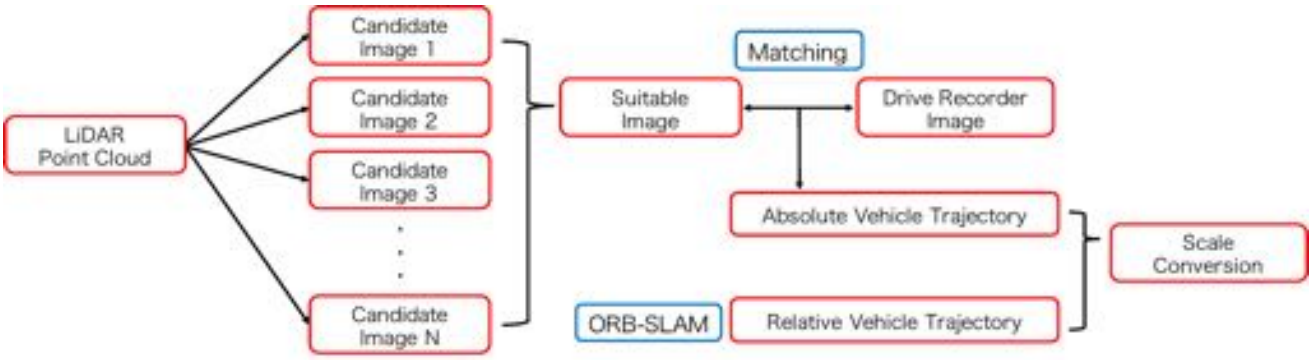

Figure 1. *Flow of the proposed method.*

Figure 2. *3D-LiDAR (Z+F IMAGER 5010C, 3D Laser Scanner).*

Figure 3. *Left: Image with distortion removed. Right: Original image.*

Figure 4. *Generated image from the LiDAR point cloud.*

Figure 5. *Matching between the drive recorder image and generated image. Left: Image generated from the LiDAR point cloud. Right: Drive recorder image with distortion removed.*

First, the previous processing is applied to the first frame of the key frame estimated by ORB-SLAM. By this processing, we can obtain the position and orientation of the vehicle of the first frame in the LiDAR point cloud. As shown in Eq. (5), the rotation axis of the ORB-SLAM estimation result is unified to the LiDAR point cloud:

$$\boldsymbol{R}_{estimated\_i} = \boldsymbol{R}_{ORB\_i}\boldsymbol{R}_{LiDAR\_init}. \tag{5}$$

Finally, we convert the relative scale trajectory to a LiDAR scale one. To calculate the scale, it is necessary to estimate the position and orientation at the absolute scale of at least two key frames. We describe the case of scale conversion by focusing on $\boldsymbol{n}$th frame. The processing of the previous section is also applied to the drive recorder image of the $\boldsymbol{n}$th frame. We calculate the scale by dividing the distance between two keyframes on the absolute scale by the output result from ORB-SLAM(see Eq. (6)):

$$scale = \frac{|\boldsymbol{t}_{LiDAR\_init} - \boldsymbol{t}_{LiDAR\_n}|}{|\boldsymbol{R}_{estimated\_init}\boldsymbol{t}_{ORB\_init} - \boldsymbol{R}_{estimated\_n}\boldsymbol{t}_{ORB\_n}|}. \tag{6}$$

We use this scale to convert the relative vehicle's trajectory to the LiDAR point cloud scale by considering the position information on the absolute scale of the initial keyframe(see Eq. (7)):

$$\boldsymbol{t}_{estimated\_i} = scale \cdot \boldsymbol{R}_{LiDAR\_i}\boldsymbol{t}_{ORB\_i} + \boldsymbol{t}_{LiDAR\_init}. \tag{7}$$

If the scale in the direction perpendicular to the road surface is scaled with the same weight as the traveling direction of the vehicle, the error becomes large, so we calculate the scale by using only the component in the traveling direction.

At this point, we would like to summarize our proposed method. First, we employ stereo ORB-SLAM, as presented by Mur-Artal et al. [13], for estimating the relative vehicle's trajectory. Next, we match the drive recorder images and generated

images from the LiDAR point cloud to calculate the relationship between the 3D LiDAR point cloud and the 2D drive recorder images. Finally, we reconstruct the vehicle's trajectory converting the ORB-SLAM estimation result by using the absolute scale estimation.

Evaluation Experiment

The experimental environment is as follows: CPU: Intel Core i7-5820K 3.30GHz, RAM: 64GB, drive recorder: KENWOOD DRV-610, LiDAR: Z+F IMAGER 5010C 3D Laser Scanner. In the evaluation experiment, we apply the proposed method to images from the drive recorder mounted on the vehicle. Then, by comparing the estimated vehicle's trajectory with the highly accurate RTKGPS, we confirm the effectiveness of our proposed method.

Input

The input images are 200 drive recorder images mounted on a vehicle traveling about 40 m (see Fig. (6)). We eliminate the distortion of these drive recorder images and input them to ORB-SLAM. The LiDAR point cloud used for scale conversion is a set of point clouds measured by setting the LidAR at 13 places at regular intervals on the road (see Fig. (7)).

Result

Figure. (8) shows the vehicle's trajectory estimated by our proposed method and the RTK-GPS which can measure position with high accuracy. The vertical axis represents the position of the vehicle in the LiDAR point cloud. Since the coordinate of the point cloud in the direction perpendicular to the ground is the Y axis, in the calculation of the scale of Eq. (6), only the X axis and the Z axis were used. This was done because the amount of change on the Y axis is overwhelmingly smaller than that on the other coordinate axes, and it is sensitive to scale conversion. The number of key frames obtained as a result of inputting 200 drive recorder images to ORB-SLAM was 20 frames. In the evaluation, we verified the accuracy by comparing the results of the key frame scale conversion with those from RTK-GPS. Figure. (8) shows the vehicle's trajectories estimated by our proposed method and RTK-GPS. The vehicle position is estimated accurately within 1.3 m of error. Figure. (9) shows the estimation error. In this experiment, only two of the obtained key frames were used for scale conversion, so errors accumulated.

Conclusion

We proposed a novel method to reconstruct the vehicle's trajectory by matching the point cloud obtained from LiDAR and drive recorder images. We input the drive recorder images to ORB-SLAM, acquired the relative track of the vehicle, and then converted them to the LiDAR scale by matching the images generated from the point cloud with the drive recorder images. In the evaluation experiments, we confirmed the effectiveness of our proposed method by comparing the position of the vehicle estimated by the proposed method with the highly accurate RTK-GPS. We showed that the vehicle's trajectory can be estimated accurately using drive recorder images and 3D-LiDAR. Finally, we can propose two points to be improved as future works. One is the process of generating images from the point cloud measured by 3D-LiDAR. Since the accuracy of our proposed method depends on matching, it is necessary to develop a rendering system that can generate a more accurate image. The other is process of scale conversion. Since scale is converted once, the estimated error tends to accumulate. We think that the accumlation of estimated error can be prevented by increasing the number of scale conversion think that the scale, we must generate images from the point cloud and this process involves a huge computational cost. Thus, we can say that there is a relationship between the accuracy and computational cost of the method.

Acknowledgments

This research presentation is supported in part by and a research assistantship of a Grant-in-Aid to the Program for Leading Graduate School for “ Science for Development of Super Mature Society ” from the Ministry of Education, Culture, Sport, Science, and Technology in Japan and ”Smart mobility system R & D and demonstration project (development of accident database construction technology)” from the Ministry of Economy, Trade and Industry in Japan.

References

[1] Pablo F Alcantarilla and T Solutions. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Patt. Anal. Mach. Intell*, Vol. 34, No. 7, pp. 1281–1298, 2011.

[2] Pablo Fernández Alcantarilla, Adrien Bartoli, and Andrew J Davison. Kaze features. In *European Conference on Computer Vision*, pp. 214–227. Springer, 2012.

[3] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, Vol. 110, No. 3, pp. 346–359, 2008.

[4] Alberto Broggi, Pietro Cerri, Stefano Ghidoni, Paolo Grisleri, and Ho Gi Jung. A new approach to urban pedestrian detection for automatic braking. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 10, No. 4, pp. 594–605, 2009.

[5] Tim Caselitz, Bastian Steder, Michael Ruhnke, and Wolfram Burgard. Monocular camera localization in 3d lidar maps. In *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*, pp. 1926–1931. IEEE, 2016.

[6] Erik Coelingh, Andreas Eidehall, and Mattias Bengtsson. Collision warning with full auto brake and pedestrian detection-a practical example of automatic emergency braking. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, pp. 155–160. IEEE, 2010.

[7] Masahiro Doi, Kai Zeng, Takahiro Wada, Shun'ichi Doi, Naohiko Tsuru, Kazuyoshi Isaji, and Shou Morikawa. Steering-assist control system on curved road using car-to-car communication. In *Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on*, pp. 1–6. IEEE, 2013.

[8] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, Vol. 24, No. 6, pp. 381–395, 1981.

[9] David Geronimo, Antonio M Lopez, Angel D Sappa, and Thorsten Graf. Survey of pedestrian detection for advanced

Figure 6. *Input images.*

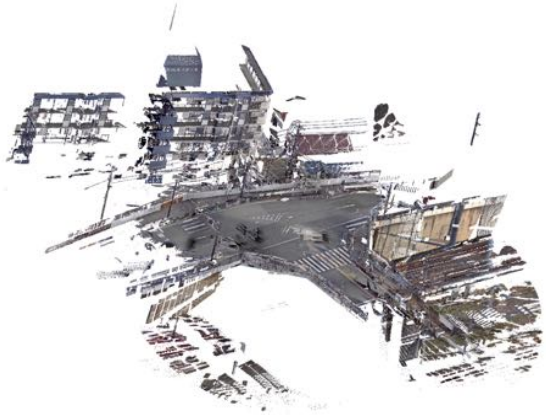

Figure 7. *Point cloud measured by 3D LiDAR.*

driver assistance systems. *IEEE transactions on pattern analysis and machine intelligence*, Vol. 32, No. 7, pp. 1239–1258, 2010.

[10] Georg Klein and David Murray. Parallel tracking and mapping for small ar workspaces. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pp. 225–234. IEEE, 2007.

[11] Jing Li, Hong Bao, Xiangmin Han, Feng Pan, Weiguo Pan, Feifei Zhang, and Di Wang. Real-time self-driving car navigation and obstacle avoidance using mobile 3d laser scanner and gnss. *Multimedia Tools and Applications*, Vol. 76, No. 21, pp. 23017–23039, 2017.

[12] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, Vol. 60, No. 2, pp. 91–110, 2004.

[13] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. Orb-slam: a versatile and accurate monocular slam system. *IEEE Transactions on Robotics*, Vol. 31, No. 5, pp. 1147–1163, 2015.

[14] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *Computer Vision (ICCV), 2011 IEEE international conference on*, pp. 2564–2571. IEEE, 2011.

[15] Kazuhiro Sakiyama, Toshihiro Shimizu, and Kazuya Sako. Vehicle driving support system, and steering angle detection device, May 20 2003. US Patent 6,567,726.

[16] Ryota Sasaki and Seiji Yasunobu. An intelligent auto-driving system by interactive acquisition of driving knowledge as information on route. In *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*, pp. 179–182. IEEE, 2004.

[17] Bo Tang, Stanley Chien, Zhi Huang, and Yaobin Chen. Pedestrian protection using the integration of v2v and the pedestrian automatic emergency braking system. In *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*, pp. 2213–2218. IEEE, 2016.

[18] Ryan W Wolcott and Ryan M Eustice. Visual localization within lidar maps for automated urban driving. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pp. 176–183. IEEE, 2014.

[19] Ziyu Zhang, Sanja Fidler, and Raquel Urtasun. Instance-level segmentation for autonomous driving with deep densely connected mrfs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 669–677, 2016.

Author Biography

Akiyoshi Kurobe received his B.E. degree in information and computer science from Keio University, Japan, in 2017. Since 2017, he has been a master student in science and technology at Keio University, Japan. His research interests include SLAM, autonomous driving, and computer vision.

Hisashi Kinoshita is employed at DENSO Corporation.

Prof. Hideo Saito received his Ph.D. degree in Electrical Engineering from Keio University, Japan, in 1992. Since then, he has been on the Faculty of Science and Technology, Keio University. In 1997 to 1999, he had joined into Virtualized Reality Project in the Robotics Institute, Carnegie Mellon University as a visiting researcher. Since 2006, he has been a full Professor of Department of Information and Computer Science, Keio University. His recent activities for academic conferences includes a Program Chair of ACCV2014, a General Chair of ISMAR2015, and a Program Chair of ISMAR2016. His research interests include computer vision and pattern recognition, and their applications to virtual reality, and human robotics interaction.

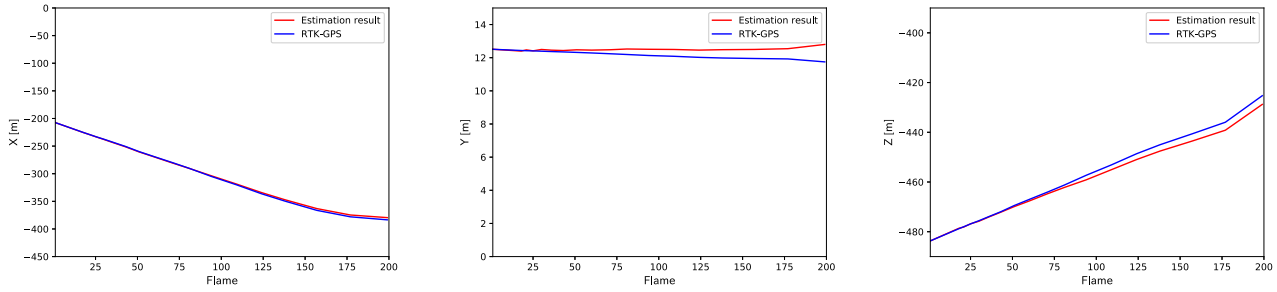

Figure 8. *Vehicle position estimation result and RTKGPS(groundtruth).*

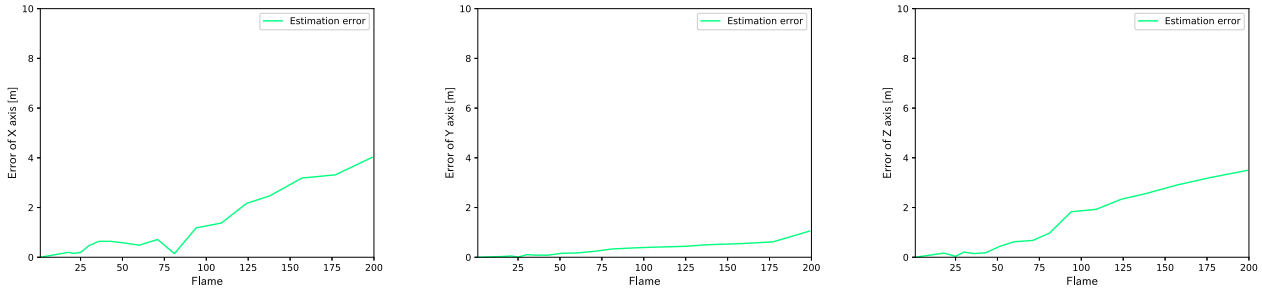

Figure 9. *Estimation error.*

Figure 10. *Left: Original drive recorder images, Center: Image with distortion removed, Right: Images generated using the estimated position and orientation.*