Vehicle Localization System in Urban Scenario with Three-Dimensional City Map

Jiali Bao 1) Yanlei Gu 1) Shunsuke Kamiyo 1)
1) The University of Tokyo, Institute of Industrial Science
4-6-1 Komaba, Meguro-ku 153-8505, Japan (E-mail: bao@kmj.iis.u-tokyo.ac.jp)

Received on September 5, 2017

ABSTRACT: Accurate self-localization is a critical problem in the autonomous driving system. In this paper, we proposed a localization system which integrates stereo camera and three-dimensional city model containing building and road mark information. The stereo camera generates visual odometry, reconstructs building scene and detects road mark. We aligned building scene with Normal Distribution Transform (NDT) map to get the absolute localization. Road mark detection result helps to rectify the inner lane positioning with road map. The experiment conducted in Hitotsubashi, Tokyo indicates that the lateral and heading error of visual odometry can be corrected and sub-meter accuracy localization is achieved.

KEY WORDS: Electronics and control, Image processing, autonomous driving system/ NDT, stereo camera, city map

1. Introduction

Accurate vehicle localization is an important but challenging part in both autonomous driving system and Advanced Driving Assistant System (ADAS) in urban scenarios. The most widely applied GNSS achieves 0.3m accuracy in open sky environment (1), but in urban scenarios, the positioning accuracy is lower because the surrounding tall buildings block the satellite signals and cause Non-Line-Of-Sight (NLOS) propagation and multipath effects (2). The error of GNSS thus increases to several meters in the urban environment. The sensor fusion which integration of GNSS and other sensors as inertial sensors helps to increase the accuracy. 3D map increases the accuracy of localization, but it is not satisfactory for autonomous driving (3, 4, 5).

Vision based method proves to be effective for vehicle localization without GNSS. Simultaneous Localization and Mapping (SLAM) technique generates the map of environment and localizes the vehicle in environment at the same time (6). Stereo vision based visual odometry system is also capable of positioning the vehicle through calculating the motion of vehicle from image sequence to track the position of vehicle (7, 8). The SLAM and visual odometry has achieved improvement in the past years, but as they are relative positioning result from stand alone system, the cumulative error of trajectory is unavoidable and tends to get greater in long distance, which will finally cause difficulties as lane identification problem. This kind of problem can be improved by the pre-prepared map as reference (9).

Following the trend to use pre-prepared map for localization, different institutes and companies are contributing to the reconstruction of maps. GNSS, Light Detection and Ranging (LiDAR) and aerial photo are used to generate 3D map (10, 11, 12). Stereo camera helps to reconstruct feature map and grid map (13). One popular format of 3D map is point cloud for localization. However, the excessive data amount of point cloud makes it difficult to be stored and utilized for real-time application. Besides, in the point cloud map, semantic information as building and road are not given.

Moreover, the use of geometric primitives to represent a large 3D map has been demonstrated as a feasible means of real-time localization (14). 3D map containing pole-like features are used to localize the vehicle (15). The road marks including traffic line information and symbolic road markings are used for localization in complicated complex urban environments (16). NDT map is a probability map that eases the computation of matching of point cloud map (17). Road map mark is also used with stereo camera for localization (18). The weak points of existing maps are discussed in (5). In 2D maps, the road marks along might be covered by other vehicles or damaged by bad weather as snow. In 3D maps, the land marks as trees and lamps are sparse and unstable in bad condition. The building information is one of the most stable and constant feature in the urban city. Therefore, in our work, we propose to adopt a Three-Dimensional City Map, which contains 3D building and 2D road information, in order to suppress the cumulative error of visual odometry and realize the accurate localization. The 3D building map, it contains only the wall position information of buildings, so the size of the map is very small and easy to access. In the road map mark, the lane information is provided by the start and end points of the line segment.

The flowchart of the proposed integrated localization system is shown in Fig.1. We pre-prepared the NDT map as reference by converting 3D building map. In the localization, only the initial position is given by GNSS. From stereo camera, we extract the building information and exclude the obstacles. With the translation and rotation from visual odometry information for several frames, the visible buildings in these time sequences are translated into accumulated observation, and then matched with NDT map. In addition, road mark is detected from time sequences and detection result is applied to match with 2D road map. After that, both 3D building matching and road mark matching results are used to estimate the current vehicle state within particle filter.

Copyright © 2018 Society of Automotive Engineers of Japan, Inc. All rights reserved
Libviso2 is used for providing visual odometry which propagates the particles. In the particle filter, the accumulated building observation is utilized to rectify the accumulated error with the aid of building model. The inner lane positioning is further decided by road mark detection and 2D road map.

The main contribution of this proposal is using the city map containing building information and the 2D road mark map, to localize the vehicle in urban scenario. Moreover, we conducted the experiments in Tokyo city center where locates a lot of tall buildings. We compared the proposed method with conventional visual odometry method and GNSS. The results of the experiment prove that the accumulated localization error and heading error are rectified.

The remaining paper is organized as follows: We introduced accumulated building observation reconstruction and 2D road mark detection in Chapter 2. The NDT map generation from 3D map is shown in Chapter 3. In Chapter 4, we discussed the localizing system which implements particle filters integrating visual odometry, NDT map and 2D map. The localization performance evaluation in Hitotsubashi is introduced in Chapter 5. Finally, we concluded this proposal with future work in Chapter 6.

2. Scene Understanding from Stereo Camera

In this paper, the building scene is reconstructed by the stereo camera system set on top of the vehicle. The flowchart of stereo image based scene understanding and local building map reconstruction is illustrated in Fig. 2. Besides the building information, we also detect the road marks from the images.

2.1. Building reconstruction and building extraction

Stereo image pairs are input into DispNet to generate the disparity map. Fig. 3 shows the image from the left camera and the corresponding disparity map in one of the intersections in our testing area. After the generation of disparity, U-disparity map is estimated. In the traffic scene, we suppose that the building is behind all the objects, therefore we filter the U-disparity map to exclude other obstacles as trees and stationed vehicle.

Firstly, we generate the U-disparity map. The U-disparity map is the histogram of disparity map in the vertical direction. In U-disparity map, every line represents a vertical object. As shown in Fig. 4(a), the obstacles as trees and traffic sign poles are compressed into lines marked in bottom rectangles and the building is represented by the pixels inside the upper rectangle.

Secondly, for the better building estimation, we need to extract the building information and exclude the obstacles. The building is supposed to be the last layer of the scene. Inspired by the U-disparity is optimized in order to exclude obstacles. Foreground elimination is applied to delete the objects of the closer layer. The obstacle could damage the continuity of building pixel in U-disparity map when the obstacle covers too much of building, thus dynamic programming is applied to cope this problem. The filtering U-disparity map result is shown in Fig. 4(b).
2.2 Generation of building observation

The filtered U-disparity map is projected in X-Z space to reconstruct the building contour as shown in Fig. 5(a). The view of camera in each frame is limited, and therefore the sequence of building reconstruction of 1 second (15 frames in our solution) is accumulated to form a complete building observation as shown in Fig. 5(b). The translation and rotation from visual odometry result are retrieved to transform the building observation. For every point in the detected building \( B_k = \{x'_i, y'_i\} \) in frame \( k \), it is converted to the observation building point \( B'_k = \{x_i, y_i\} \) by

\[
\begin{bmatrix}
x'_i \\
y'_i
\end{bmatrix} = 
\begin{bmatrix}
\cos \theta_k & -\sin \theta_k & T_{k,x} \\
\sin \theta_k & \cos \theta_k & T_{k,y}
\end{bmatrix}
\begin{bmatrix}
x_i \\
y_i
\end{bmatrix}
\]

(1)

where \( i \) is the point index, \( T_{k,x} \), and \( T_{k,y} \) are the translation vectors and \( \theta_k \) is the deviation angle obtained from visual odometry relative to the first frame in the sequence.

In this proposal, we applied road mark detection to localize the vehicle inside lane. The detection algorithm is as follows:

1) Free space detection: In order to exclude the obstacle and increase the accuracy of road mark detection, free space is first detected via Stixel [21]. In Fig.6(a), the free space is marked green and only the pixels in the free space will be taken into line detection.

2) Inverse Perspective Mapping (IPM): The top view of road scene, IPM, is generated before road mark detection. In the IPM, the road marks become parallel and the pixel represents the lateral physical distance, which eases the road mark detection and lateral positioning.

3) Image filtering: The road mark information is presented as vertical lines in IPM, thus the horizontal lines are filtered to decrease noise and increase detection accuracy. The filtering result is shown in Fig. 6(c).

4) Hough transformation: The Hough transformation is applied on the filtered image to get Region of Interest (ROI) containing the line pixels, which is marked as a rectangle in Fig. 6(d).

5) Ransac line detection: Ransac is applied on the ROI of Hough transformation detection to get a fine road mark detection in Fig. 6(e). The final multi-lane detection result is shown in Fig. 6(f).

3. Normal Distribution Transformation Map Generation

In this proposal, NDT map is used to match with building reconstruction result. The NDT map describes the probability that a point is observed in a location. We pre-prepared the NDT map to align the accumulated building observation and weight the particle. The cell size is set to be 5m. The NDT map is generated by the following steps:

1) Point sampling of building map: The concise building map is composed of lines of the contours of buildings, so the lines are sampled to be points with the resolution of 10 cm.

2) Cell division: The map space is divided into cells with constant size. In this solution, the size of one cell is set to be 5m by 5m.

3) Mean and Covariance matrix calculation for one NDT cell: All the occupied points \( x_i \) inside each cell are collected. The mean \( q \) and covariance matrix \( \Sigma \) of all the occupied points are calculated by equation (2) and (3).

\[
q = \frac{1}{n} \sum x_i
\]

(2)

\[
\Sigma = \frac{1}{n} \sum (x_i - q)(x_i - q)^T
\]

(3)

where \( x_i \) is the point sampled inside the cell and \( n \) is the number of points. The Fig.7 shows the NDT map generated from building map. Fig. 7(a) shows the original building contours which are extracted from the 3D data of Hitotsubashi area. The wall information is compacted and saved as lines. Fig. 7(b) illustrates the NDT map generated from building contours. In the NDT map, lighter color means higher probability and darker color represents lower probability.

4. Self-Localization with Particle Filter
In the self-localization system, we implemented particle filter to integrate the city map with building and road mark detection result. In this proposal, we utilize the visual odometry generated by libviso2 for particle propagation (11). Particle filter utilizes a set of particles \( P_k = \{ P_{k,i} \}^n \), where \( P_{k,i} \) is a 3 dimension coordinate of each particle, \( i \) is the particle index, \( k \) indicates the kth epoch of driving, \( N \) and \( E \) denote localization in the north and east direction, \( \Theta \) is the heading direction and \( n \) is the amount of particles in the set. In addition, each particle has its own weight to indicate the importance of particle. The weight of each particle is \( w_{k,i} \). The position and weight of particle are estimated by the following steps:

4.1. Prediction

Based on the previous particles \( \{ P_{k-1,i}^{e}, w_{k-1,i}^{e} \}^n \) of epoch \( k-1 \), the current particles \( \{ P_k^i, w_k^i \}^n \) of epoch \( k \) are calculated through the motion of the vehicle as shown in (4) and (5).

\[
P_k = \begin{bmatrix}
P_{k,x}^i \\
P_{k,y}^i \\
P_{k,z}^i \\
N_k^i \\
E_k^i \\
\Theta_k^i
\end{bmatrix} = \begin{bmatrix}
P_{k,i-1,x}^{e} \\
P_{k,i-1,y}^{e} \\
P_{k,i-1,z}^{e} \\
N_k^{i-1} \\
E_k^{i-1} \\
\Theta_k^{i-1}
\end{bmatrix} + \begin{bmatrix}
\Delta N_k^i \\
\Delta E_k^i \\
\Delta \Theta_k^i
\end{bmatrix}
\]  

\[
\begin{bmatrix}
\Delta N_k^i \\
\Delta E_k^i
\end{bmatrix} = \begin{bmatrix}
\sin(P_{k,i}^x) & \cos(P_{k,i}^x) \\
-\cos(P_{k,i}^x) & \sin(P_{k,i}^x)
\end{bmatrix} \begin{bmatrix}
T_x \\
T_y
\end{bmatrix} 
\]

where \( \Delta N_k \) and \( \Delta E_k \) are the translation vector in north and east direction, and \( \Delta \theta \) is the deviation angle obtained from visual odometry, and \( T_x \) and \( T_y \) represent the lateral and longitudinal translation from visual odometry.

4.2. Weight evaluation

The particle is weighted by both the matching local building with building map and road mark detection with 2D map. The process of weight evaluation by building map is shown in Fig. 8. In Fig. 8, the points represent the particles, the lines on the particles are the directions of particles and the depicted pixels on the top-right corner are the local building reconstruction result. The score of every particle by matching local building reconstruction and NDT map is calculated by

\[
score(B \mid P^i_k, M) = \sum_{j=1}^{LP} \exp \left\{-\left(\frac{x_{i,j}^k - q^j}{\Sigma_n^{-1}}\right)^2\right\}
\]

where \( B \) is the accumulated observation, \( M \) is the NDT map, \( x_{i,j}^k \) is the local building points converted into particle view. \( q^j \) and \( \Sigma_n \) are the mean and covariance of cell in NDT map where the point \( x_{i,j} \) locates. The probability of particle is thus

\[
p(B \mid P^i_k, M) = \frac{score(B \mid P^i_k, M)}{L_num}
\]

where \( L_num \) is the number of points in local building map \( B \). In Fig. 8, the probability of every particle is shown by different grey scale. The darker color is lower probability and the lighter color represents higher probability. The lines from the particles represent the heading direction of every particle.

Road mark detection and 2D map also contribute to the particle weight. The result of weight evaluation for one particle by road mark detection is shown in Fig. 9. In Fig. 9, the point inside the circle represents one particle. The two arrowed lines indicate the distance from the particle to the left road line \( D_{lane,left}^{i,k} \) and to right road line \( D_{lane,right}^{i,k} \). The probability \( p(x, \{ P^i_k \} \) of road mark detection result for given the particle \( P^i_k \) is calculated by equation (8)-(10) as follows:

Fig. 7 3D building map (a) and NDT map (b) particles in the set. The lines are contours of buildings.
The estimated position and direction is finally calculated by the weighted average of the particle positions:

$$P_k^{pf} = \sum w_i^t P_k^i$$

4.3. Resampling:

Particles $\{P_k^i, w_i^t\}_{i=1}^n$ are resampled based on their weights. The particles of low weight are replaced with new particles around the center of high weight particles.

5. Experiment

To evaluate the accuracy of localization, we conducted the experiments in the Hitotsubashi area of Tokyo city where the density of tall building is high. The accurate city map is one of the main resources of the system. For building map, we employed a company to provide 3D data by measuring the experiment area with Mobile Mapping System (MMS) car. From the 3D data, we generate 3D building map. In the 3D data, the building is represented by the position and height of each corner, and they are converted to extract the boundary information. The 2D road map is created based on the aerial image. Fig. 11 depicts the city map containing building information and road information. To evaluate the system in different condition, the experiment area includes various traffic scene as one-way roads, intersections and two-way roads.

For the visual equipment, we built the flexible high-resolution long-baseline stereo system platform as shown in Fig. 12. The camera platform is installed on the carrier of on top of the vehicle. The stereo camera consists of two cameras with the resolution of 2048x1500. As the position of camera can be changed on the platform, the baseline of the stereo camera can be set from 30cm to 90cm for different scenarios. In the urban scenario, the buildings to be detected can be as far as 50 meters away from the vehicle. To estimate these buildings, we chose the baseline of 80 cm. These images from the stereo camera sets are the input of building detection, visual odometry and road mark detection. Also, from these images, we manually distinguished the ground truth of localization.

In vehicle self-localization, the lateral position is more important as the lateral position decides in which lane the vehicle is driving. Fig. 13 defines the lateral position error and heading error. The lateral error $Error_k$ of position $P_k$ is the minimal distance from the position to the ground truth trajectory. The heading error is the angle difference between the ground truth trajectory and estimated trajectory.

In the experiment in Hitotsubashi, we ran 4 drives and the total distance is approximately 4 kilometers. Fig. 14 shows the trajectory and one drive localization result example. To evaluate the localization accuracy, we compared our proposed method with conventional visual odometry and GNSS. In Fig. 14, the lines are the ground truth. The squares represent GNSS results by GNSS receiver on top of our vehicle. The triangles are conventional visual odometry trajectory given by libviso2 and the circles are the result of the proposed method. From the localization result in Fig. 14, we...
can see that the accumulated error of visual odometry is rectified by our proposed method.

The quantitative performance of localization is summarized in Table 1. We evaluated the mean and standard deviation of localization of conventional visual odometry, GNSS and our proposed method for four drives. Also, we compared the heading direction errors of visual odometry and our proposed method. Our proposed method provides 0.84m Lateral error and 0.031 rad heading error on average. The conventional visual odometry method gives 1.85m Lateral error and 0.070 rad heading error, and GNSS provides 2.62m lateral error in the same condition on average. We achieved sub-meter localization accuracy in all four drives. As the width of the driving lane is about 3.5m in Japan, our proposed localization system accurately localizes the vehicle in the correct lane.

To give a more direct impression of performance comparison, Fig. 15 illustrated the error of one drive. In Fig. 15, the yellow line is the performance of the conventional method and the red line represents our proposed method. In the most epochs, our proposed method achieved localization performance of lower than 1 meter errors. However, for some specific epoch, the error increases to about 2 meters. The reason for these errors is that in some specific intersection, where the building is of very far distance from the vehicle and building estimation accuracy is low. When the vehicle comes closer to the building, the error decreases again.
Table 1 Localization performance in Hitotsubashi

(mean (standard deviation))

| Proposed method | Conventional method | GNSS |
|-----------------|---------------------|------|
| Drive           | Lateral Error (m)   | Heading Error (rad) | Lateral Error (m) | Heading Error (rad) |
|                 | (0.76)              | (0.026)          | (2.06)            | (0.080)            | (2.02) |
| 1               | 0.83                | 0.024            | 3.32              | 0.034              | 3.03  |
|                 | (0.51)              | (0.080)          | (0.79)            | (0.055)            | (1.72) |
| 2               | 0.81                | 0.038            | 1.07              | 0.059              | 1.87  |
|                 | (0.71)              | (0.025)          | (0.84)            | (0.086)            | (2.21) |
| 3               | 0.81                | 0.022            | 1.27              | 0.027              | 2.53  |
|                 | (0.79)              | (0.081)          | (1.83)            | (0.082)            | (2.53) |
| 4               | 0.92                | 0.039            | 1.73              | 0.121              | 3.05  |
|                 | (0.79)              | (0.081)          | (1.83)            | (0.082)            | (2.53) |
| Total           | 0.84                | 0.031            | 1.85              | 0.071              | 2.62  |
|                 | (0.70)              | (0.053)          | (1.38)            | (0.076)            | (2.12) |

6. Conclusion

In this paper, we proposed to use stereo camera and 3D city map to realize the accurate localization. In the proposed system, building and road marking are detected from stereo vision system. Those detection results are aligned with a prepared map for localization with a particle filter. The experiment in Tokyo indicates that sub-meter accuracy can be achieved by this solution. The proposed system can rectify the cumulative error of stereo visual odometry with the aid of building map and 2D road map to achieve accurate vehicle localization in the urban environment. In the future work, we will focus on motion planning based on the localization and scene understanding.

Acknowledgement

The presented research was supported by the Leading Graduates Schools Program, “Global Leader Program for Social Design and Management of The University of Tokyo” by the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

(1) L. T. Hsu, Y. Gu, S. Kamijo: NLOS Correction/Exclusion for GNSS Measurement Using RAIM and City Building Models, Sensors, Vol. 15, No. 7, pp. 17329-17349 (2015)
(2) G. MacGougan, G. Lachapelle, R. Klukas, K. Siu, L. Garin, J. Shewfelt, G. Cox: Performance Analysis of a Stand-alone High-sensitivity Receiver, GPS Solutions, Vol.6, No.3, pp. 179-195 (2002)
(3) L. T. Hsu, Y. Gu, S. Kamijo: 3D Building Model-based Pedestrian Positioning Method using GPS/GLONASS/QZSS and Its Reliability Calculation, GPS solutions, Vol. 20, No. 3, pp. 413-428 (2016)
(4) Y. Gu, L. T. Hsu, S. Kamijo: Passive Sensor Integration for Vehicle Self-Localization in Urban Traffic Environment, Sensors, Vol. 15, No. 12, pp. 30199-30220, (2015)
(5) S. Kamijo, Y. Gu, L. Hsu: Autonomous Vehicle Technologies: Localization and Mapping, IEICE Fundamentals Review, Vol. 9, No. 2, pp. 131-141 (2015)
(6) J. Engel, T. Schöps, D. Cremers: LSD-SLAM: Large-scale direct monocular SLAM, Proceedings of European Conference on Computer Vision, pp. 834-849 (2015)
(7) H. Badino, T. Kanade: A Head-Wearable Short-Baseline Stereo System for the Simultaneous Estimation of Structure and Motion, Proceedings of IEEE Conf on Machine Vision Applications., pp. 185-189 (2011)
(8) A. Geiger, J. Ziegler, C. Stiller: StereoScan: Dense 3d Reconstruction in Real-time, Proceedings of IEEE Conf on Intelligent Vehicles Symposium, pp. 963-968 (2011)
(9) M. A. Brubaker, A. Geiger, R. Urtasun: Map-based Probabilistic Visual Self-localization, IEEE Trans. Pattern analysis and machine intelligence, Vol. 38, No.4, pp. 652-665 (2016)
(10) Y. Wada, L.-T. Hsu, Y. Gu, S. Kamijo: Optimization of 3D Building Models by GPS Measurements, GPS Solutions, pp. 1-14 (2015)
(11) A. Hornung, K. M. Wurm, M. Bennwewitz, C. Stachniss, W. Burgard: OctoMap: An Efficient Probabilistic 3D Mapping Framework Based on Octrees, Autonomous Robots, Vol.34, No.3, pp. 189-206 (2013)
(12) L. Zebedin, A. Klaau, B. Gruber-Greymayer, K. Karner: Towards 3D Map Generation from Digital Aerial Images, ISPRS Journal of Photogrammetry and Remote Sensing, Vol.60, No.6, pp. 413-427 (2006)
(13) Y. Li, Y. Ruichel: Occupancy Grid Mapping in Urban Environments from a Moving On-board Stereo- vision System, Sensors, Vol.14, No.6, pp. 10454-10478 (2014)
(14) B. Soheilian, O. Tournaire, N. Paparoditis, B. Vallet, J. P. Papelard: Generation of an Integrated 3d City Model with Visual Landmarks for Autonomous Navigation in Dense Urban Areas, Proceedings of IEEE Conf on Intelligent Vehicles Symposium, pp. 304-309 (2011)
(15) R. Spangenberg, D. Goehring, R. Rojas: Pole-based Localization for Autonomous Vehicles in Urban Scenarios, Proceedings of IEEE Conf on Intelligent Robots and Systems, pp. 2161-2166 (2016)
(16) J. K. Suhr, J. Jang, D. Min, H. G. Jung: Sensor Fusion-Based Low-Cost Vehicle Localization System for Complex Urban Environments, IEEE Trans. Intell. Transp. Syst., Vol. 18, No. 5, pp. 1078-1086 (2016)
(17) P. Biber, W. Straßer: The Normal Distributions Transform: A New Approach to Laser Scan Matching, Proceedings of Intelligent Robots and Systems, IEEE/RSJ International Con. on, Vol. 3, pp. 2743-2748 (2003)
(18) T. Senlet, A. Elgammal: A framework for Global Vehicle Localization using Stereo Images and Satellite and Road Maps, Proceedings of Computer Vision Workshops, IEEE Conf on, pp. 2034-2041 (2011)
(19) N. Mayer, E. Ilg, P. Hausser, P. Fischer, D. Cremers, A. Dosovitskiy, T. Brox: A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation, Proceedings of Conference on CVPR, pp. 4040-4048 (2016)
(20) Z. Hu, F. Lamosa, K. Uchimura: A Complete Uv-disparity Study for Stereo Vision Based 3d Driving Environment Analysis, Proceedings of IEEE Conf on 3-D Digital Imaging and Modeling, pp. 204-211 (2005)
(21) H. Badino, U. Franke, D. Pfeiffer: The Stixel World-A Compact Medium Level Representation of the 3d-world, Proceedings of Joint Pattern Recognition Symposium, pp. 51-60 (2009)