

Visual Simultaneous Localization and Mapping (vSLAM) of Driverless Car in GPS-Denied Areas [†]

Abira Kanwal, Zunaira Anjum and Wasif Muhammad * 

Intelligent Systems Laboratory, Department of Electrical Engineering & Technology, University of Gujrat, Gujrat 50700, Pakistan; abirakanwal577@gmail.com (A.K.); zuneeraanjum.456@gmail.com (Z.A.)

* Correspondence: syed.wasif@uog.edu.pk

[†] 1st International Conference on Energy, Power and Environment, University of Gujrat, Gujrat, Pakistan, 11–12 November 2021.

Abstract: A simultaneous localization and mapping (SLAM) algorithm allows a mobile robot or a driverless car to determine its location in an unknown and dynamic environment where it is placed, and simultaneously allows it to build a consistent map of that environment. Driverless cars are becoming an emerging reality from science fiction, but there is still too much required for the development of technological breakthroughs for their control, guidance, safety, and health related issues. One existing problem which is required to be addressed is SLAM of driverless car in GPS denied-areas, i.e., congested urban areas with large buildings where GPS signals are weak as a result of congested infrastructure. Due to poor reception of GPS signals in these areas, there is an immense need to localize and route driverless car using onboard sensory modalities, e.g., LIDAR, RADAR, etc., without being dependent on GPS information for its navigation and control. The driverless car SLAM using LIDAR and RADAR involves costly sensors, which appears to be a limitation of this approach. To overcome these limitations, in this article we propose a visual information-based SLAM (vSLAM) algorithm for GPS-denied areas using a cheap video camera. As a front-end process, features-based monocular visual odometry (VO) on grayscale input image frames is performed. Random Sample Consensus (RANSAC) refinement and global pose estimation is performed as a back-end process. The results obtained from the proposed approach demonstrate 95% accuracy with a maximum mean error of 4.98.

Keywords: localization; mapping; visual odometry; pose estimation



Citation: Kanwal, A.; Anjum, Z.; Muhammad, W. Visual Simultaneous Localization and Mapping (vSLAM) of Driverless Car in GPS-Denied Areas. *Eng. Proc.* **2021**, *12*, 49. <https://doi.org/10.3390/engproc2021012049>

Academic Editor: Qasim Awais

Published: 29 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Simultaneous localization and mapping (SLAM) is a problem in which a mobile robot or driverless car is placed in an unknown environment, and has to determine its location and simultaneously build a consistent map. All previous approaches are computationally expensive and memory-space hungry, which limits them to develop a dynamic map for a limited area.

A technique based on Symmetries and Perturbation (SP) mapping was applied to solve the problem of estimating a vehicle's current location and forming a map, but it failed when a robot navigated through some previously visited places [1]. In [2], Monte Carlo localization was used to counter for the global location approximation problem. A vision-based approach was used to detect and track the location of dynamic objects, but it worked inefficiently in cases with less sensory information [3]. For navigation in very large areas, a map merging technique for multi-robot SLAM was applied in conjunction with reinforcement learning [4]. The geometric features of the environment were tracked using ultrasonic sensors, but this approach had high computational complexity [5]. Landmark based navigation was used along with a feature extraction algorithm, but it was computationally insufficient [6]. A SLAM problem was also addressed with an unscented Kalman

filter to solve the problems of computational complexity and memory constraints [7]. A technique based on adaptive place networks was used to enable a robot to navigate robustly, but this approach was not efficient [8].

To overcome these limitations, in this article we developed a monocular vision-based SLAM algorithm for GPS-denied locations. With a monocular SLAM approach, a single camera was used as it is more computationally efficient, cost effective, and less memory hungry. The main advantages of monocular visual SLAM are the low cost and it is easy to implement [9].

2. Methods

Image sequences captured from different congested environments by a monocular camera, mounted at the top of a car, provide visual information to perform vSLAM by visual odometry in four steps, i.e., features detection, features tracking, RANSAC refinement, and global pose estimation. As corners are the most salient part of any object, corners are chosen as invariant features in images and detected using Features from an Accelerated Segment Test (FAST) detector. The detected features are tracked by applying Sparse Optical Flow (SOF) using the Kanade–Lucas (KL) method. Flow vectors are formed by determining a new positions (u,v) of tracked points in the present frame using the KL method.

An essential matrix is determined by using epipolar geometry, which contains rotation and translation information. The epipolar constraint used in this technique is $y_2^T E y_1 = 0$. Here, y_1 and y_2 are homogenous normalized image coordinates. The results of features tracking are not perfect and may have a great deal of incorrect correspondence; therefore, Nister's 5-point algorithm [10] is employed with RANSAC to find an essential matrix. The intrinsic camera parameters (focal length and optical center) are used to obtain the camera matrix. In the next step, a pose is estimated by finding camera rotation and translation by using an estimated essential matrix and the consistent feature points in two consecutive images. The essential matrix is defined as $E = [t]_{\times} R$. Here, R is the rotation matrix and t is the translation matrix, and $[t]_{\times}$ is the matrix illustration of a cross product with t . It adopts a chirality check to find relative camera rotation and translation and the inliers which permit that check are returned.

An estimated essential matrix, E , can be decomposed using Single Value Decomposition (SVD) as illustrated in Equation (1) and then utilizing the constraints to find the translation and rotation matrices by using Equations (2) and (3):

$$E = U \Sigma V^T \quad (1)$$

$$[t]_{\times} = U W \Sigma U^T \quad (2)$$

$$R = U W^{-1} V^T \quad (3)$$

For a given E , there are four possible poses: $[R1, t]$, $[R1, -t]$, $[R2, t]$, and $[R2, -t]$. The chirality check recovers the pose by findings those 3D triangulated points which have positive depth. The global pose is estimated by using the ground truth information of the trajectory. It is achieved by finding the Euclidean distance between the previous ground truth pose and the current ground truth pose for every current frame and then multiplying the recovered pose with its estimated global pose. As a final step, a map is constructed by concatenating the estimated poses, which are both rotational poses and translational poses as given in Equations (4) and (5).

$$t_{pos} = t_{pos} + tR_{pos} * scale \quad (4)$$

$$R_{pos} = RR_{pos} \quad (5)$$

The rotational and translational poses are initialized with zero, and for every next frame translation and rotation matrices are obtained which contain the information of rotational and translational distance covered by the vehicle from previous frame to the

next frame and the location of vehicle in that frame. For every frame, the rotational and translational matrices (pose) are updated and concatenated to develop the map.

3. Results

The experiments were performed in outdoor congested urban areas with large buildings in the surroundings of the road, and to collect the dataset the input was taken from a camera. All experiments conducted in this article for the evaluation of proposed SLAM algorithm KITTI grayscale datasets [11], containing 4540 image sequences, were used. As the first process the corner features were detected from the surroundings. The parameters of feature detection algorithm were set such that it provided approximately 4000 features from a single image, as shown in Figure 1a. The features detection step was carried out for all the images present in the dataset and features were detected according to the environment. These features were tracked based on detected points from the next frame, and motion vectors were drawn by the optical flow method on image frames with colorful lines. The tracking results for the first and the second frames are shown in Figure 1b.

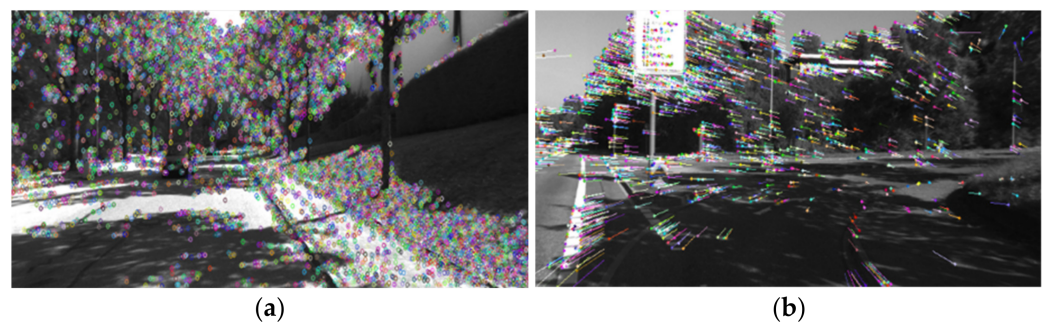


Figure 1. (a) Detected features in an image frame; (b) motion vectors of tracked features.

The poses that were obtained from global pose estimation are (x,y) coordinates of estimated location of a vehicle. The corresponding ground truth poses with (x,y) coordinates from the KITTI dataset were used to compare the estimated vehicle poses and ground truth poses for the first 200 frames of the KITTI dataset, as shown in Figure 2a. The estimated poses are depicted with blue dots and the ground truth is represented with red triangles. In addition, the trajectory of the vehicle by concatenating global estimated poses relative to ground truth poses are compared and shown in Figure 2b. In this paper, eight different datasets were used to obtain simulation results of vSLAM. The map was constructed using 500 image frames from each of the eight datasets.

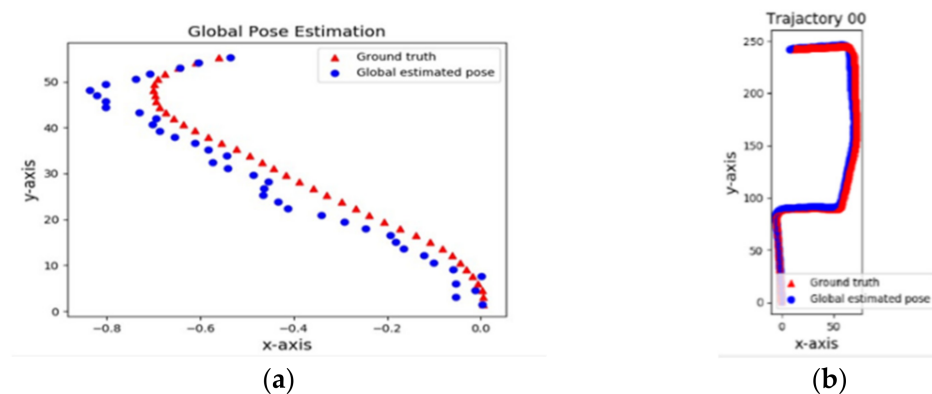


Figure 2. (a) Ground truth and estimated poses for first 200 frames; (b) estimated and ground truth trajectory.

There is a small scale drift error between ground truth and global estimated poses at certain locations in maps. Poses were concatenated to build the map, due to which very small errors that were present in every global estimated pose accumulated, resulting in scale drift. The result of RMS error plots for dataset 00 trajectory are shown in Figure 3a, and mean RMS errors for all eight datasets are shown in Figure 3b. It can be seen that the maximum mean error obtained was 4.98 in the case of dataset 07. This amount of error is negligible, as SLAM is a run-time process so small amounts of drift can be ignored. As a result of the proposed SLAM algorithm, both localization and mapping simultaneously are realized for a driverless car in congested urban environment.

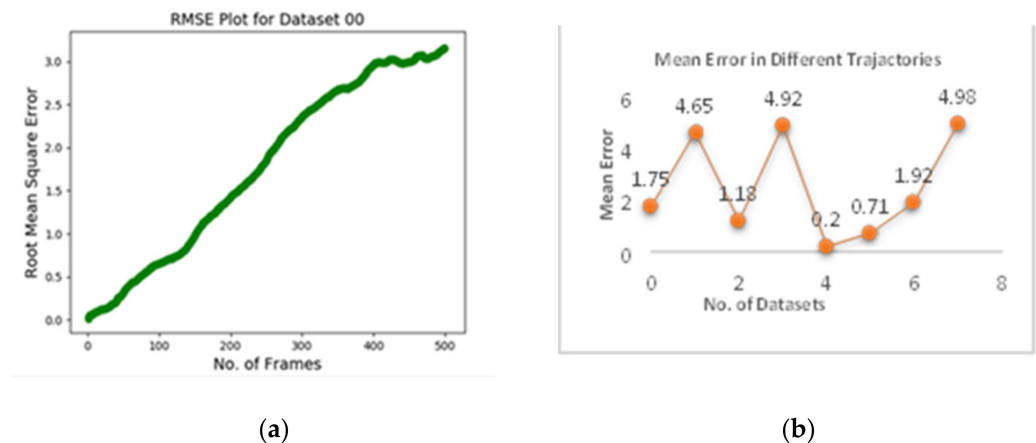


Figure 3. (a) Root mean square error plot for 00 dataset; (b) root mean square error for 8 datasets.

4. Conclusions

In this article, we proposed a vSLAM algorithm for driverless car navigation in GPS-denied areas. The localization was performed using visual odometry, in which fast corners as features were detected from a grayscale image and then these features were tracked in successive frames using an optical flow method. The motion of a driverless car was estimated by finding the essential matrix and then a local pose was recovered by finding rotation and translation matrices and by using the essential matrix. The maximum encountered mean rms error was 4.98 due to scale drift. However, as SLAM is a real-time process this error can be ignored. Visual odometry with a cheap monocular camera offered a cheaper and mechanically easier method for motion estimation.

Data Availability Statement: The data used in this article is a publicly available data from this source; http://www.cvlibs.net/datasets/kitti/raw_data.php (accessed on 1 September 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Castellanos, J.A.; Montiel, J.M.M.; Neira, J.; Tardos, J.D. The SPmap: A probabilistic framework for simultaneous localization and map building. *IEEE Trans. Robot. Autom.* **1999**, *15*, 948–952. [\[CrossRef\]](#)
2. Dellaert, F.; Fox, D.; Burgard, W.; Thrun, S. Monte Carlo localization for mobile robots. In Proceedings of the 1999 IEEE International Conference on Robotics and Automation (Cat. No. 99CH36288C), Detroit, MI, USA, 10–15 May 1999; Volume 2.
3. Dellaert, F.; Burgard, W.; Fox, D.; Thrun, S. Using the condensation algorithm for robust, vision-based mobile robot localization. In Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No pr00149), Fort Collins, CO, USA, 23–25 June 1999; Volume 2.
4. Dinnissen, P.; Givigi, S.N.; Schwartz, H.M. Map merging of multi-robot slam using reinforcement learning. In Proceedings of the 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Seoul, Korea, 14–17 October 2012.
5. Leonard, J.J.; Hugh, F.D.-W. Simultaneous map building and localization for an autonomous mobile robot. *IROS* **1991**, *3*, 1442–1447.
6. Lerner, R.; Rivlin, E.; Shimshoni, I. Landmark selection for task-oriented navigation. *IEEE Trans. Robot.* **2007**, *23*, 494–505. [\[CrossRef\]](#)

7. Wen, S.; Chen, X.; Ma, C.; Lam, H.K.; Hua, S. The Q-learning obstacle avoidance algorithm based on EKF-SLAM for NAO autonomous walking under unknown environments. *Robot. Auton. Syst.* **2015**, *72*, 29–36. [[CrossRef](#)]
8. Yamauchi, B.; Beer, R. Spatial learning for navigation in dynamic environments. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **1996**, *26*, 496–505. [[CrossRef](#)] [[PubMed](#)]
9. He, M.; Zhu, C.; Huang, Q.; Ren, B.; Liu, J. A review of monocular visual odometry. *Vis. Comput.* **2020**, *36*, 1053–1065. [[CrossRef](#)]
10. Yousif, K.; Bab-Hadiashar, A.; Hoseinnezhad, R. An overview to visual odometry and visual SLAM: Applications to mobile robotics. *Intell. Ind. Syst.* **2015**, *1*, 289–311. [[CrossRef](#)]
11. Geiger, A.; Lenz, P.; Urtasun, R. Are we ready for autonomous driving? The kitti vision benchmark suite. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012.