Real-time vehicle localization using on-board visual SLAM for detection and tracking

Nabil Belbachir
NORCE Research
nabe@norceresearch.no

Nadia Saad Noori
NORCE Research
nano@norceresearch.no

Benyamin Akdemir
NORCE Research
bakd@norceresearch.no

Paper ID 7

Abstract

This paper presents an on-board system based using a monocular vision system, auxiliary sensors and a visual SLAM method for real-time vehicle localization and tracking in natural outdoor environments. A monocular panamorph camera providing 360° panoramic views has been integrated with a gyro, compass and accelerometer to form a sensor system for installation on-board vehicles. A direct method was applied for detection and tracking of environment based on the images and combined with a calibration of the keypoints using the auxiliary sensors for providing an accurate trajectory to the vehicle. An-board computer is used to facilitate a standalone deployment on-board vehicle. The system was developed and designed as an integral component of a security monitoring and tracking system for humanitarian missions and convoys operating in regions with poor or absence of GPS signal coverage. Preliminary results show the applicability of the system with acceptable accuracy compared to GPS after evaluation in natural outdoor environments in Europe.

1. Introduction

Recent years are witnessing rapid and intense activities on concepts, sensors, algorithms and software development for autonomous localization and tracking in natural environment. This is driven by the large interest and needs within autonomous navigation without GPS for indoor mobile applications and for the self-driving autonomous cars sectors. The real-time visual Simultaneous Localization and Mapping (SLAM) [1][2][3][4][5][6] has gained higher attention and larger contributions from the scientific community, making use of images as input data with the objective of estimating the camera trajectory while reconstructing the visual environment. This reconstruction is very relevant for application in robotics and autonomous systems with cameras are mounted on-board mobile platforms (e.g. smartphones, vehicles, aircrafts and drones). Two categories of visual SLAM methods are typically used with monocular cameras. The first category, called feature-based visual SLAM methods [1][2] that typically detect and match a set of visual features across a sequence of images using descriptors extracted from every image frame. As most of the information contained in each image is discarded, matching sparse key points leads to efficient bundle adjustment and computation at the cost of reduced accuracy and robustness. The second category, called direct methods [6][7], makes use of the full image to directly minimize the photometric error between an image and a map for increased accuracy, dense reconstructions and some robustness to viewpoint change and blur. Direct methods can use and reconstruct the whole such that decision is based on more complete information compared to feature-based method where only a fraction of information is used. The main limitation for these methods is the implicit assumption of static scene illumination required for the photometric error metric, which is only valid in controlled environments (e.g. indoor) and not in natural dynamically changing environments. Direct methods have challenges to remove outliers retroactively and needs to have a good initialization.

In this research, however, we have selected LIBVISO2, a feature-based method [8] that resembles a direct method in producing a semi-dense 3D representation of the mapped area. However, LIBVISO2 uses a second CPU thread, it only performs these calculations at 1-2 fps, while feature extraction and analysis happens at more than ten times the rate of the 3D reconstruction. The method produced better results in evaluation for localization and mapping in natural outdoor environments. To reduce visual SLAM limitations mentioned above, an auxiliary sensor box has been used to continuously (re)initiate the visual SLAM for avoiding cumulative errors and minimize the outliers. The sensor box includes an accelerometer, a compass and a gyro in order to facilitate scaling the trajectory from the image space into the world coordinates. Within this research, we were interested in apply this method on 360° panoramic views from a camera system equipped with a panamorph lens [9][10] as provides minimal distortion. We evaluate the combination of this camera system mounted on front a vehicle (Fig. 1) for generating wide views, direct method for a dense reconstruction and auxiliary sensor for error minimization and scaling the trajectory into the world coordinates. The paper is structured as follows: In section 2, a brief description of the sensor system is given. Section 3 presents visual SLAM method used. The feasibility analysis of the method on images generated by a panamorph camera for real-time deployment given in section 4 including the calibration approach. Section 5 provides a short concluding summary.

Fig. 1. Illustration of panoramic imaging in front of vehicle (credit [11]).

2. System description

The system has been designed to operate on-board vehicle for autonomous localization and tracking in unconstrained environment. The main task to research and develop on-board sensing with capabilities for vehicle localization (in absence of GPS) with constraints towards achieving a functional technology but also affordable in cost and deployment in practice. The system has to meet the following framework conditions for acceptability within the targeted security market:

1. It has to be low-cost to be competitive for this market.
2. It has to be self-contained for easy deployment and use on-board vehicles.
3. It has to be easy-mountable and un-mountable to allow non-specialists for deployment in the field.
4. It does not have to rely on any equipment in the car (e.g. odometer, gyros) apart the power. This is because either vehicles do not offer interface possibility (manufacturers do not grant interface) or old vehicles do not offer digital interfaces.

Fig. 2. illustrates a car equipped with the sensor system. The camera mounted in front of the car makes use of a panomorph lense [11] developed by ImmerVision. It is an advanced hemispheric wide-angle lens and is designed with patented anamorphosis (optimal sensor coverage) and/or magnification (targeted distortion), to enable covering more pixels and magnify desired zone of interest by controlling the distortion through optical design. The optical distortion is used to magnify zones of interest and to increase resolution, which enables to distribute the pixel density along the detector photosensitive surface. The concept anamorphosis provides optimal pixel coverage where the image is being stretched to optimize coverage on a rectangular sensor. The comparative analysis of the panamorph lens and fish-eye lens of the resolution gain shows the following[12]:

1. Using a standard fisheye lens on a VGA imager (640×480 pixels), the resulting resolution is a constant over the entire field-of-view and can be calculated as follows:

$$R_{fisheye} = 480 \text{ pixels} / 190 \text{ deg} = 2.52 \text{ pixels} / \text{deg} \qquad (1)$$

2. With a panomorph lens, the image mapping can be managed in a manner that the resolution for each zone will be the same, and we can decrease the resolution in the inter-zone (between the forward and side views). This approach is claimed to provide a better control over the resolution. The panomorph lens includes an anamorphic correction, hence the 190 degree field-of-view will be spread not only over 480 pixels but on the longer axis, and will use 640 pixels (30% more). If we consider that the inter-zone also covers 38 degrees and that the resolution can be reduced by a factor two in this zone, the new resolution of the panomorph lens is:

$$R_{panamorph} = 160 \text{ pixels} / 38 \text{ deg} = 4.21 \text{ pixels} / \text{deg} \qquad (2)$$

3. In this calculation, it is considered that 160 pixels are required to image each zone of interest and 80 pixels are required for each inter-zone for a total of 640 pixels. The resolution provided by the panomorph lens seems to be 170% higher than the resolution provided by a conventional fisheye lens in the zones of interest. For the inter-zones, the resolution is only 10% smaller.

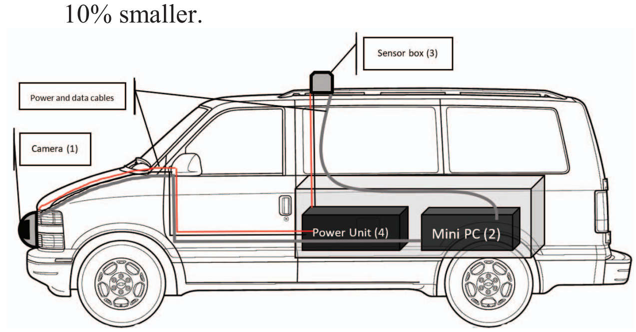

Fig. 2. Illustration of the system mounted on-board vehicle

The camera is supported by sensor box components (Fig.3) integrating new sensors such as a compass (geomagnetic sensor), accelerometer and gyro.

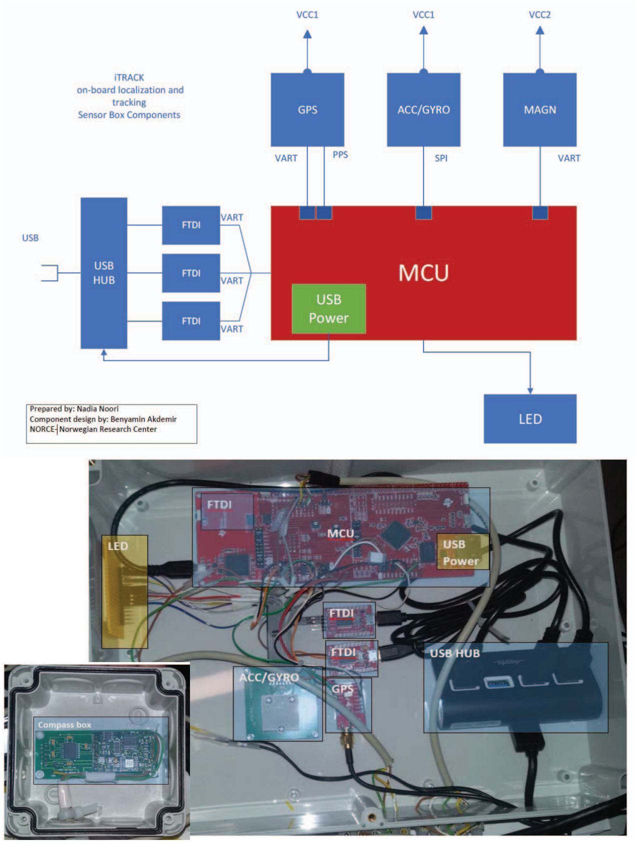

Fig. 3. Picture of the sensor box components

On-board Mini PC is used for real-time processing. We have combined the sensor data information with the image analysis output for a robust tracking. Sensor box inputs are incorporated (See Fig. 4) into the image analysis algorithm to increase the accuracy of the localization and to correct for scale and orientation issues as a cheap alternative to 3D vision using depth imaging systems. The output from the on-board computer is the 2D position of the camera, and thereby the object/vehicle the camera is attached to, relative to a reference point, updated every 10 second and sent to the to the onboard mobile device through Bluetooth communication.

3. SLAM method- LIBVISO2 library

LIBVISO2 extract features in a greedy fashion, which provided a less complex process than other methods (e.g. ORB-SLAM2) for detecting points [13]. As a fast feature-based visual odometry (VO) library for monocular and stereo cameras [14], LIBVISO2 provided the capability to have real-time tracking for our system. Similar to other feature-based methods, it consists of feature matching over subsequent frames and egomotion estimation by minimizing the reprojection error. Features are extracted by filtering the images with a corner and blob mask and performing non-maximum and non-minimum suppression on the filtered images. Starting from all feature detections in the current left image, candidates are matched in a circular fashion over the previous left image, the previous right image, the current right image, and back to the current left image. If the first and last features of such a circle match differ, the match is rejected. Based on all found matches, the egomotion is then estimated by minimizing the reprojection error using Gauss-Newton and outliers are removed using RANSAC [8] .

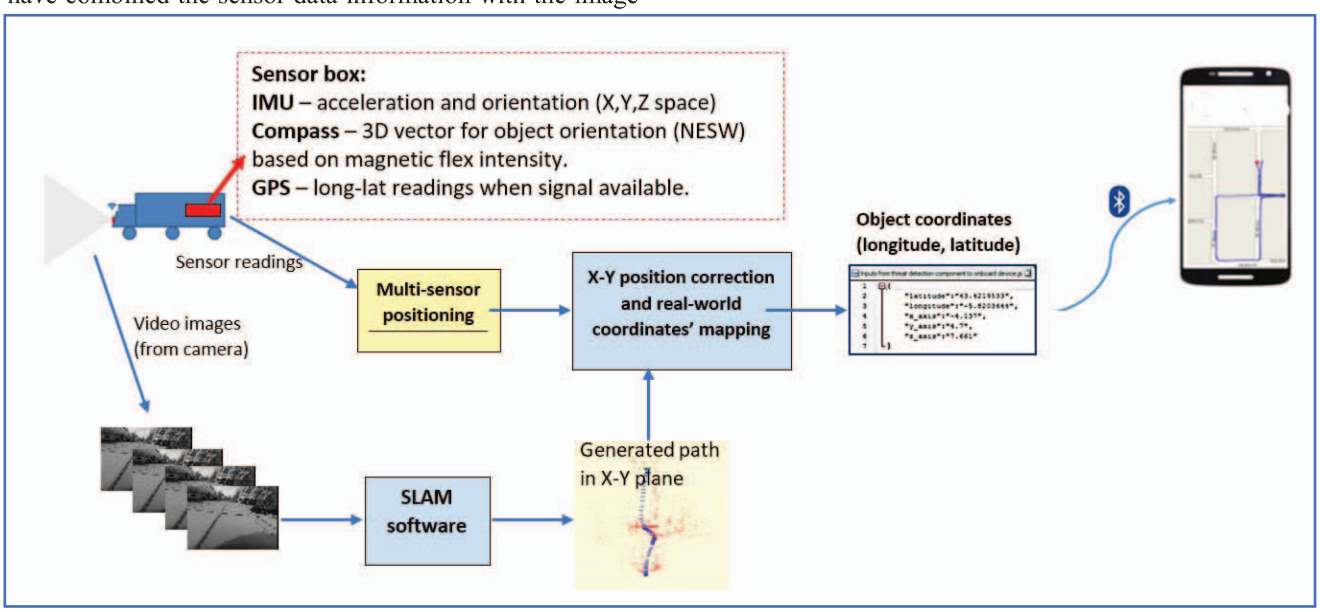

Fig. 4. Illustration of the system information flow

4. Feasibility analysis of the system

For the evaluation of this system few runs were performed. The first run was dedicated for data collection, which was performed in Norway. The other runs were made in Norway and the Netherlands for testing and optimizing of the system and reduction of the resulting vehicle trajectory error compared to the one resulting from GPS.

The used camera has a 6 megapixels resolution generating images at 25 frames per second. A high-performance mini PC is attached to the system to register and process data and estimate the vehicle location on the image views and its scaling into the world coordinates. The mini-PC is a Zotac Magnus PC with a quad core of 3.6 GHz

For the evaluation of the accuracy, the raw GPS position (converted into x and y image coordinates) and was incorporated into the algorithm to compare the location deviation with respect to the visual SLAM location results.

The vehicle was driving at an average speed of 50 km/h and the trajectory using both GPS and SLAM is depicted on x-y space in Fig. 5. We can notice two right turns at 450m and 1100 m and one left turn at 700m. At 1300m, there was a roundabout showing a direction change on the figure below. We can notice that the SLAM trajectory is intersecting with the GPS trajectory resulting in few zero errors. This is due to the regular initialization triggered by the sensor box to minimize the error drift and keep the accuracy under control.

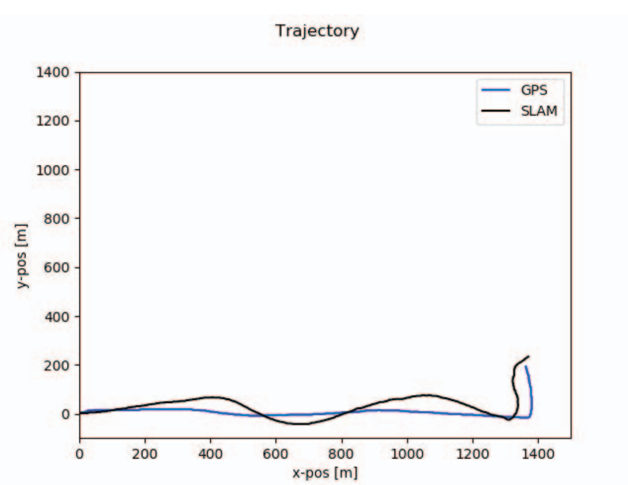

Fig. 5. Results of the vehicle trajectory from the SLAM and GPS on x-y representation.

TABLE I depicts a quantitative analysis of accuracy of the visual SLAM localization for every few meters where the position given by the direct method is regularly corrected using the sensor box information and is compared with the position given by the GPS. Quantitative Results of the Trajectory Accuracy compared to that of GPS. The results presented in TABLE I are based on data collected during the final evaluation session in Netherlands, however, previous evaluation sessions during the design phase of the system had provided similar results.

Distance (m)	Error (m)
10	4
100	3
200	15
300	31
400	55
500	48
600	14
700	38
800	16
900	16
1000	47
1100	71
1200	46
1300	7
1400	42
1500	72
1600	51
1630	41

TABLE I. Error calculation between SLAM and GPS results along 1630m

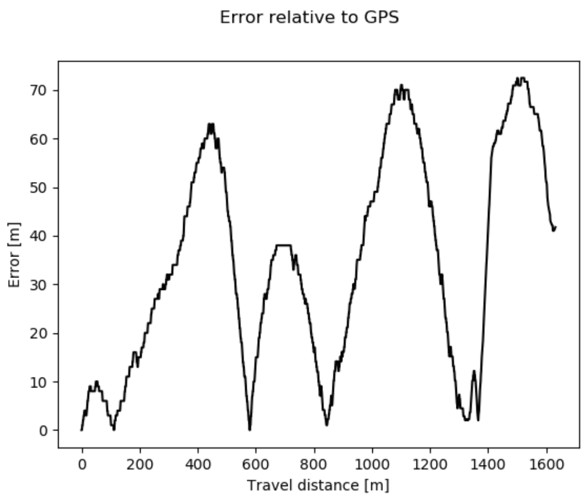

Fig. 6. Error visualization between SLAM and GPS results along 1630m.

5. Conclusions and Outlook

In this paper, two main messages are transmitted. First, with this work, we showed that it is possible to perform real-time vehicle localization using a direct method on monocular 360° panoramic views based on a non-conventional wide-view lens (panamorph). Panamorph lenses are claimed to provide single camera panoramic views at higher optical resolution than that of fish-eye lenses using the same imager chip. Secondly, we showed that the integration of a sensor box including an accelerometer, gyro and a compass were able to reduce the direct method disadvantages and regularly re-initiate the location to hinder large drifts and keep the error under control for long distance. This system has potential to be deployed for localization and navigation in unconstrained outdoor environment in an easy manner. This system provides new localization capabilities, which can be exploited in robotics and autonomous driving applications at low-cost deployment. This work is a feasibility analysis, which shows the applicability of the combination of both panamorphic cameras with a visual SLAM and auxiliary sensors, but still further testing is required to validate the results and further optimize the accuracy for longer trajectories.

References

[1] R. Mur-Artal, J. M. M. Montiel and J. D. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147-1163, October 2015.

[2] H. Lim, J. Lim, and H. J. Kim, "Real-time 6-DOF Monocular Visual SLAM in a Large-scale Environment," in IEEE International Conference on Robotics and Automation (ICRA), pp. 1532–1539, Hong Kong, China, June 2014.

[3] E. Eade and T. Drummond, "Scalable monocular SLAM," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 469–476, New York City, USA, June 2006.

[4] G. Pascoe, W. Maddern, M. Tanner, P. Pinies and P. Newman, "NID-SLAM: Robust Monocular SLAM Using Normalised Information Distance," In Proc. of IEEE CVPR, pp. 1436-1444, USA, 2017.

[5] K. Tateno, F. Tombari, I. Laina and N. Navab, "CNN-SLAM: Real-Time Dense Monocular SLAM With Learned Depth Prediction," In Proc. of IEEE CVPR, pp. 6243-6252, USA, 2017.

[6] G. Siveira, E. Malis and P. Rives, "An Efficient Direct Approach to Visual SLAM," IEEE Transactions on Robotics, vol. 24, issue 5, pp. 969-979, October 2008.

[7] J. Engel, V. Koltun, and D. Cremers. "Direct Sparse Odometry". In: IEEE Transactions on Pattern Analysis and Machine Intelligence (Mar. 2018).

[8] A. Geiger, J. Ziegler, and C. Stiller, "Stereo Scan: Dense 3D Reconstruction in Real-time", In: Intelligent Vehicles Symposium (IV). 2011.

[9] S. Thibault, J. Parent, M. Larivière-Bastien, A.S. Poulin-Girard, A. Arfaoui and P. Desulniers, "Developments in modern panoramic lenses: lens design, controlled distortion, and characterization," in SPIE Proceeding, vol. 8197, Beijing, China, 2011.

[10] S. Thibault, A. Arfaoui and P. Deslauniers, "Cross-diffractive optical elements for wide angle geometric camera calibration," In Optics letters, 36, pp. 4770-4772, 2011.

[11] S. Thibault, "Novel compact panomorph lens based vision system for monitoring around a vehicle," in Optical Sensors, vol. 7003, 2008.

[12] S. Thibault, "360° Vision System: Opportunities in Transportation," Proceedings of SPIE, the International Society for Optical Engineering, Society of Photo-Optical Instrumentation Engineers, 67580H. 1-67580H. 9., 2007.

[13] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Visual Odometry / SLAM Evaluation2012. url: http://www.cvlibs.net/datasets/kitti/eval_odometry.php.

[14] N. Krombach, D. Droeschel, S. Houben, and S. Behnke, "Feature-based visual odometry prior for real-time semi-dense stereo SLAM," *Robotics and Autonomous Systems,* vol.109: 38-58, 2018.