Journal of Computer Science

Review

Comparative Study of Visual Odometry Performance Based on Road Classifications

Dayang Nur Salmi Dharmiza, Awang Salleh and Kuryati Kipli

Department of Electrical and Electronics, Faculty of Engineering, University Malaysia Sarawak, Malaysia

Article history
Received: 27-07-2022
Revised: 26-08-2022
Accepted: 04-10-2022

Corresponding Author:
Dayang Nur Salmi Dharmiza
Department of Electrical and Electronics, Faculty of Engineering, University Malaysia Sarawak, Malaysia
Email: asdnsdharmiza@unimas.my

Abstract: Accuracy and robustness are among the main concerns in vehicle positioning systems and autonomous applications. These concerns are crucial in GNSS-denied environments; thus, we need an alternative technology to overcome this problem. In recent years, vision-based localization known as visual odometry has gained considerable attention among researchers. Visual odometry is a vision-based pose estimation and it has been developed for mobile object localization such as robots and vehicles while perceiving their environment. Within the last decade, researchers have been immersed in developing techniques to achieve highly accurate and precise localization based on visual odometry. The visual odometry performances are evaluated using an online dataset for benchmarking. Based on the benchmarking, this study reviews and compares the robustness of the recent visual odometry techniques for application, especially in vehicle localization in various road conditions. Evaluation methods for the selected techniques are presented and a thorough analysis of each driving sequence is conducted. The analysis shows that for all visual odometry techniques, localization for high-speed drive suffers higher translation error even though the surrounding has less image noise. Despite that, visual odometry that implements careful feature Selection and Tracking (SOFT) proves to be more robust compared with other techniques with 0.7% relative translation error and a relative rotation error of 0.2 deg/hm.

Keywords: Visual Odometry, Localization, Autonomous Vehicle

Introduction

With the rapid technological advancement in the field of mobile robotics and automation, growing demand has arisen for the accurate localization of moving objects. One of the motion estimation techniques that is gaining popularity is vision-based odometry thanks to its low cost, simplicity, and wide application of the camera itself. Besides, since cameras are robust and passive sensors, they are the leading candidates to facilitate in a GNSS-denied environment. This vision-based odometry is also known as Visual Odometry (VO).

Visual Odometry (VO) is the process of estimating the position and orientation of a mobile object by analyzing continuous camera images (Nistér *et al*., 2004). Until today, VO has been widely applied to various mobile robotic platforms, visual and augmented reality, and wearable devices (Mukhopadhyay, 2014). Especially with the prevalence of the development of autonomous or driverless vehicles, VO has become an interesting research field in computer vision and positioning systems.

However, since vehicles are driven on the road at various speeds under different weather types and environments, the robustness of VO is questionable. Indeed, with the research development, VO accuracy is optimized, however, the performance indicator of certain VO techniques is mostly based on the average positioning error of multiple sequences experimented. This positioning error is computed from the relative translation error and the positioning relative rotation error to the ground truth of the vehicle. In this study, we review the different techniques of VO systems developed in the last few years briefly and evaluate their performances according to different road types.

Related Visual Odometry Works

Generally, the VO systems can be categorized into three approaches: Feature-based, appearance-based (direct), and hybrid-based (semi-direct) systems as depicted in Fig. 1. Feature-based VO consists of two parts which are the feature management and the state optimization steps. This approach benefits from robust modern point-feature descriptors such as BRIEF, (Calonder *et al*., 2010), BRISK, (Leutenegger *et al*., 2011), ORB (Rublee *et al*., 2011), and FREAK (Alahi *et al*., 2012).

DOI: 10.3844/jcssp.2022.1030.1037

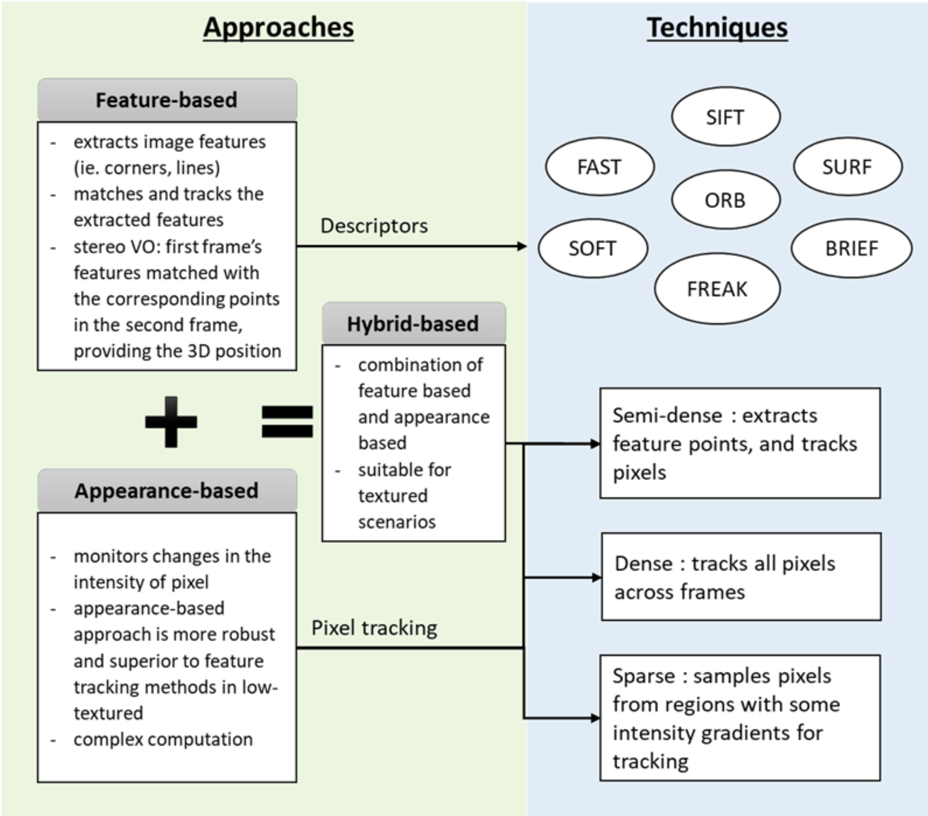

Fig. 1: Visual odometry approaches and techniques

ORB-SLAM2 proposed by Mur-Artal *et al*. (2015) is one of the most cited VO methods that utilized a feature-based approach, with ORB features for tracking, mapping and place recognition tasks, proves to be accurate and robust to motion clusters in most scenarios. Other published VO systems with a feature-based approach are presented by Geiger *et al*. (2011); Bénet and Guinamard (2020); Krešo and Šegvic (2015); Wang *et al*. (2019); Cvišić *et al*. (2018; 2022a). In their works, (Geiger *et al*., 2011; Bénet and Guinamard, 2020; Krešo and Šegvic, 2015) used corner and blob convolution such as Harris Corner detector and then employ non-maximum- and non-minimum-suppression on the filtered images (Neubeck and Gool, 2006). As for Joint Forward-Backward Visual Odometry (JFBVO) introduced by Wang *et al*. (2019), they proposed an interesting idea of a novel method with the joint forward-backward framework which incorporates cues from backward motion to improve the forward motion estimate. Meanwhile, Cvišić *et al*. (2018; 2022ab) recently developed their VO methods with SOFT feature tracking that is based on careful selection and tracking of stable features whereas the latter work optimized its accuracy based on the camera recalibration presented in Cvišić *et al*. (2022b). From their works, the SOFT feature tracking technique has shown an outstanding performance in improving VO accuracy.

However, feature-based VO tends to have high latency due to the expensive computation of data association. To solve this, appearance-based VO systems directly find the optimal geometric transformation by minimizing the photometric error between the input image and the warped reference frame. Among the noteworthy VO, systems are LSD-SLAM (Engel *et al*., 2014), Direct Sparse Odometry (DSO) (Engel *et al*., 2017), and Gradient-based Joint Direct Visual Odometry (GDVO) (Zhu, 2017). These works employed an appearance-based (direct) approach while Semi-direct Visual Odometry (SVO) as proposed by Forster *et al*. (2016) utilizes a semi-direct approach as its name implies. One of the direct approaches, DSO, for instance, implements a sparse formulation that can significantly reduce computation complexity, unlike the dense pixel tracking proposed by Meilland *et al*. (2011); Newcombe *et al*. (2011) and semi-dense pixel tracking implemented in Engel *et al*. (2014); Zhu (2017) of previous researches. This meant that DSO is capable of achieving real-time computation, as it samples only points of sufficient intensity gradient and neglects the geometric prior.

Evaluation Datasets

In parallel with the advancing development of autonomous robots and vehicles, public datasets are

DOI: 10.3844/jcssp.2022.1030.1037

essential as they enable evaluation and comparison of different approaches. As for visual odometry and Simultaneous Localization and Mapping (SLAM), several datasets have been made publicly available over the years such as the KITTI dataset (Geiger *et al*., 2013), Málaga Urban dataset (Blanco-Claraco *et al*., 2014), KITTI-360 dataset (Liao *et al*., 2022), The EuRoc micro aerial vehicle dataset (Burri *et al*., 2016), Oxford Robotics Car dataset (Maddern *et al*., 2017), Multivehicle Stereo Event Camera Dataset (MVSEC) (Zhu *et al*., 2018) and a Stereo Event Camera Dataset (DSEC) (Gehrig *et al*., 2021).

Among these, the most established and widely used for VO evaluation purposes is the KITTI dataset. The KITTI dataset contains 11 image sequences recorded from a car in urban and highway environments. The recordings total up to 40 min, but individual recording for each sequence ranges from 30 sec to 8 min. The car is equipped with several sensors: Including four cameras, a Velodyne laser scanner, and an accurate Inertial Navigation System (GPS/IMU). To validate VO performance, ground truth positions provided by RTK-GNSS are used. This study's focus is on a performance evaluation review of the KITTI dataset only because it has a large-scale outdoor benchmark that is suitable for self-driving applications. The KITTI dataset has been developed into the KITTI-360 dataset (Liao *et al*., 2022), where the driving sequence is longer and has more sensory information with both static and dynamic 3D scene elements. However, this dataset is too new and only a few have published their evaluation results on the leaderboard.

Road Classifications

As mentioned previously, the KITTI dataset has recordings of urban and highway roads for localization evaluation. The sequences are categorized into three types of roads: Residential, city, and highway. These roads have their characteristics as shown in Table 1.

The speed limit on the roads varies according to the country's traffic regulations. Since this dataset was obtained in Germany, the speed limit for residential areas is 30 km/h, city road is 50 km/h and the highway speed limit is 130 km/h. The road shape and surrounding environment are also different for each road type. For residential roads, the surroundings are mostly residential buildings like houses and apartments, with lots of trees and parked vehicles at the roadside. The roads are narrow, usually single-lane roads. Besides, there are lots of cross-junctions and T-junctions to connect the residential paths. Meanwhile, fewer junctions can be found on city roads and the road is wider with clearer lane marks.

As for highway roads, the shape is less complex to ensure safe high-speed driving. Highway roads typically consist of multiple lanes in the same direction, so the view is cleaner from the noise contributed by other moving objects. However, there are road divergences for highway exits and at highway entrance, the roads would merge. This affects the vehicle path planning if the localization is not accurate at the lane level (Awang Salleh and Seignez, 2018).

Table 1: Road classification and characteristics

Road type	Speed limit	Shape	Environment
Residential road	30 km/h	• Multiple junctions	• Static vehicles parked at road sides
		• Narrow roads	• Building shadows
City road	50 km/h	• Less junctions	
		• One-lane or two-lane roads	• Traffic lights
			• Other moving vehicles in different vehicles
Highway road	130 km/h	• Straight or slightly curved	• Other moving vehicles in the same direction
		• Road divergent or merging (highway exit and entrance)	• Fewer buildings

Table 2: Details on the 11 sequences tested for VO performance evaluation

Sequence	Raw data	Environment	Length (m)	No of frames (10fps)	Min speed (km/h)	Max speed (km/h)	Average speed (km/h)	Loop Closure
00	2011_10_03_drive_0027	Residential	374.2	4540	0	36	13.0	Yes
01	2011_10_03_drive_0042	Road (highway)	2453.2	1100	0	65	43.0	No
02	2011_10_03_drive_0034	City + residential	5067.2	4660	0	50	27.0	Yes
03	2011_09_26_drive_0067	Residential	560.9	800	NA	NA	NA	No
04	2011_09_30_drive_0016	Road	393.6	270	46	56	50.0	No
05	2011_09_30_drive_0018	Residential	2205.6	2760	0	41	13.5	Yes
06	2011_09_30_drive_0020	Residential	1232.9	1100	0	16	4.5	Yes
07	2011_09_30_drive_0027	Residential	694.7	1100	0	37	13.0	Yes
08	2011_09_30_drive_0028	Residential	3222.8	4070	0	44	18.0	No
09	2011_09_30_drive_0033	City + residential	1705.1	1590	0	50	34.0	Yes
10	2011_09_30_drive_0034	Residential	919.5	1200	0	20	4.0	No

DOI: 10.3844/jcssp.2022.1030.1037

Table 2 describes the details of each sequence. We also include the raw data name, sequence road type, length, and loop closure status in the table. We are unable to obtain the speed information for sequence 03 due to the unavailability of the raw file for sequence 2011_09_26_drive_0067 in the KITTI dataset. Therefore, the evaluation for sequence 03 is also omitted in this study.

Out of 11 sequences provided by KITTI for evaluation, nine of them are recorded in the residential area with an average speed of not more than 20 km/h. The minimum speed for all the sequences is 0 km/h due to the vehicle stopping at junctions or traffic lights, except for sequence 04 where the trajectory is generated as a short non-stop drive on a straight road. The highest speed is recorded from a drive on a highway-sequence 01-at 65 km/h. The longest drive is sequence 02 with a 5 km driving scene that includes a city road and a residential road. Of the nine sequences in the residential area, six of them contain loop closure-sequence 00, 02, 05, 06, 07, and 09. The trajectories for all sequences (except sequence 03) are illustrated in Fig. 2.

Localization Accuracy Evaluation

The accuracy of the visual odometry technique is quantified from the estimated position evaluation concerning the ground truth as shown in Fig. 3. This evaluation is necessary, especially in benchmarking the system with the existing techniques. There are several methods for measuring the accuracy of vehicle positioning techniques. So far there is no fixed indicator for accuracy, resulting in quite a several types of research having their definitions and can sometimes be misleading. However, the most popular metrics used are the Absolute Trajectory Error (ATE) and Relative Projection Error (RPE) metrics.

The ATE evaluates the global consistency of localization by comparing the absolute distances of the estimated pose with the ground truth. Therefore, the ATE can be defined as the Root Mean Square Error (RMSE) for both rotation (Eq. 1) and positioning error (Eq. 2).

$$ATE_{rot} = \left(\frac{1}{N} \sum_{i=0}^{N-1} \left\| \angle \left(\Delta R_i \right) \right\|^2 \right)^{\frac{1}{2}} \tag{1}$$

$$ATE_{pos} = \left(\frac{1}{N} \sum_{i=0}^{N-1} \left\| \left(\Delta p_i \right) \right\|^2 \right)^{\frac{1}{2}} \tag{2}$$

Here, ΔR_i is the angle error with ground truth, Δp_i is the pose error, and the $\angle$ (.) means the rotation matrix is using the angle-axis representation and the rotation angle is the error.

ATE has one advantage; it is easy to compare localization performances because it provides a single number metric for the position/rotation/velocity estimation. However, ATE can be sensitive to the time when the error occurs. For instance, a rotation estimation error tends to give a higher ATE when it occurs at the beginning of the trajectory than the situation when it occurs at the end. Therefore, the relative error method provides another option to give a more informative evaluation of the localization accuracy.

On the other hand, the RPE measures the relative relation between the states at a fixed time interval Δ. Thus, the RPE relates to the drift of the trajectory, which is useful for the evaluation of VO accuracy. Similar to the ATE, RPE is also divided into translational and rotational errors. Firstly, the relative pose error is defined as:

$$E_i := \left(Q^{-1} Q_{i+\Delta} \right)^{-1} \left(P_i^{-1} P_{i+\Delta} \right) \tag{3}$$

where, Q is the ground truth and P is the estimated pose. From a sequence of n poses, we obtain $m = n - \Delta$ as the individual RPE matrices along the sequence. Hence, the RPE can be computed as follows:

$$RPE_{rot} = \frac{1}{M} \sum_{i=0}^{M-1} \angle \left(rot \left(E_i \right) \right) \tag{4}$$

$$RPE_{pos} = \left(\frac{1}{M} \sum_{i=0}^{M-1} \left\| E_i \right\|^2 \right)^{\frac{1}{2}} \tag{5}$$

Since the RPE generates a collection of errors for all the sub-trajectories instead of a single number for the sequence, we can calculate the statistics on the median, average, and percentiles and this gives more detailed information than ATE. Besides, RPE can provide different meanings according to different criteria selection. For example, the RPE obtained from a closer interval would reflect in the local consistency, while the error for a larger distance reflects more on the long-term accuracy. For this reason, the KITTI dataset evaluation computes translational and rotational errors for all possible subsequences of length (100, ..., 800) meters which are followed by all the researchers for fair benchmarking.

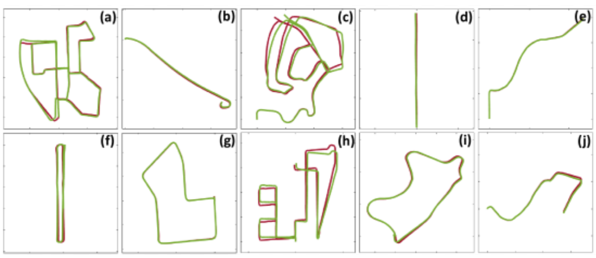

Fig. 2: Trajectories of Sequence (a) 00, (b) 01, (c)0 2, (d) 04, (e) 05, (f) 06, (g) 07, (h) 08, (i) 09, (j) 10

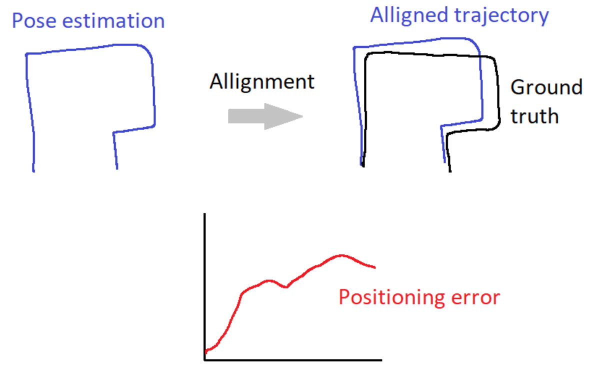

Fig. 3: The process of quantitative trajectory evaluation

Results Evaluation

This study suggests that the road types would impact the VO performance for vehicle localization because of the differences in the speed limit, road shape, and environmental effects on visual-based localization. Especially if loop detection is applied, this will re-correct the vehicle positioning and remove the accumulated errors of the positioning system. Besides, VO-based localization benefits from its lower speed, hence increasing its accuracy. This makes VO performs better on residential roads compared with other sequences.

For quantitative evaluations, the KITTI leaderboard is ranked based on the Relative Translation Error (RTE), t_{rel}, which averages the trajectory drift over segments of lengths ranging from 100 m to 800 m. When computing the performance score for the KITTI leaderboard, the average is calculated over all the segments of all sequences (not the mean of t_{rel} over the sequences) and this reflects on the leaderboard ranking. The Relative Rotation Error (RRE), r_{rel}, is also computed for all possible subsequences of length from 100 m to 800 m, thus the RRE is presented in degrees per hundred meters (deg/hm).

Since the majority of the tested sequences are recorded from the residential area, this caused a biased performance evaluation-very minimal evaluation on higher speed is conducted. Therefore, to fairly compare the performances of different VO techniques, we perform the error comparison for each sequence. We selected 13 VO techniques - VISO2 (Geiger *et al*., 2011), LSD-VO (Engel *et al*., 2014), 2FO-CC (Krešo and Šegvic, 2015), ORBSLAM2 (Mur-Artal *et al*., 2015), SOFT-SLAM (Cvišić *et al*., 2018), VINS-Fusion (2018), SOFT-VO (Cvišić and Petrović, 2015), GDVO (Zhu, 2017), StereoDSO (Wang *et al*., 2017), JFBVO (Wang *et al*., 2019), RADVO (Bénet and Guinamard, 2020), OV2-SLAM (2021), and SOFT2 (Cvišić *et al*., 2022ab)-for performance comparison. Unfortunately, RADVO did not provide its RRE for all the sequences, so we only used its average value published on KITTI's leaderboard. Besides, 2FO-CC also did not evaluate their technique on sequences 01, 02, and 03. Therefore, its performance only reflects the localization for sequence 04 until sequence 10. The details for each of the VO techniques proposed can be found in their published works.

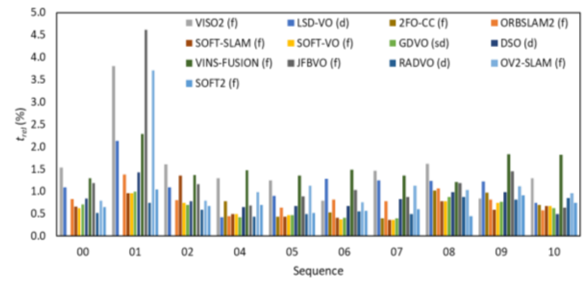

Fig. 4: RTE for each sequence

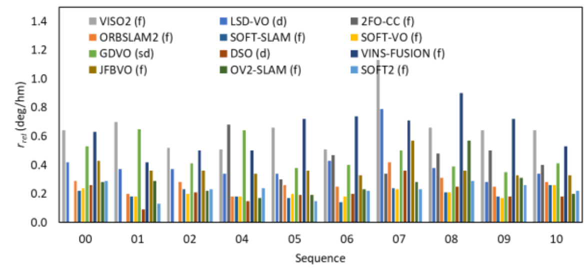

Fig. 5: RRE for each sequence

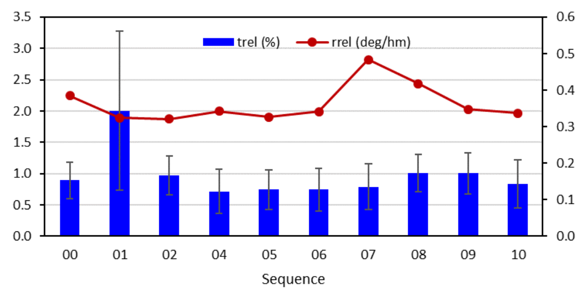

Fig. 6: RTE and RRE average for each sequence

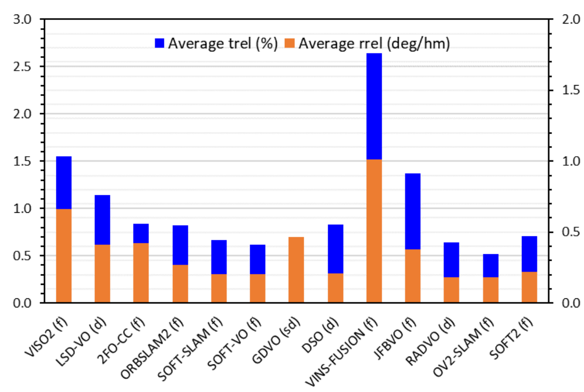

Fig. 7: RTE and RRE average for each VO technique

As can be seen from the RTE chart in Fig. 4, the highest error is recorded by JFBVO in sequence 01 (highway road) with over 4%. Sequence 01 is 2.5 km in length with a maximum speed of 65 km/h. Other VO methods also suffer from higher translation errors for this sequence compared with other sequences. Ten out of twelve VO methods (excluding 2FO-CC) recorded sequence 01 as the sequence with the highest RTE. Only SOFT-SLAM and RADVO show no distinct error whereas SOFT-SLAM recorded the highest RTE in sequence 03 at 1.36% and RADVO's highest RTE was obtained in sequence 08 at 0.88%. The high RTE of most VO methods in sequence 01 is mainly due to the drift over time which can be contributed to the scale drift or rotation error. But the most contributing factor is the high speed, which significantly affects the scale factor even with minimal projection error. This agrees with the VO performance evaluation results obtained for sequence 01.

On the other hand, as shown in Fig. 5, sequence 07 records the highest RRE by VISO2 of 1.13 deg/hm. Not only VISO2, but eight other VO methods also achieved the highest RRE for this sequence. Upon observation, this is mainly caused by an instant where the vehicle stopped at a T-junction and other vehicles are moving horizontally in front of it. This affects the rotation calculation in VO. Only SOFT-based VO techniques managed to score a rotation error of less than 0.3 deg/him and yet this is still higher compared with the rotation error of other sequences for the same VO method.

To summarize VO performances for each sequence. Fig. 6 displays the mean and standard deviation of both translational and rotation errors. Sequence 04 obtains the most precise results which are expected due to the nature of the sequence (straight, non-stop drive) with 0.71% average RTE. Sequence 01 exhibits a high RTE of 2% on average while interestingly its average rotation error is among the lowest (0.32 deg/hm) – owing to minimal noise from the environment (building/road signs/vehicle from the opposite direction) in view for VO trajectory generation. However, this still does not portray the capability of VO in a real scenario where the average speed for the highway is around 90 km/h and the drive distance is farther. Sequence 07 has the highest RRE average of 0.48 deg/hm although its average for RTE is good (0.79%).

As for the overall VO performances, we illustrate the average error in Fig. 7 in ascending order of publication date starting with VISO2 in 2011, with their approach notation -(d) for direct, (f) for feature-based, and (sd) for semi-direct. The RTE average ranges from 0.52% (OV2-SLAM) to 2.64% (VINS-Fusion) while the RRE is between 0.18 deg/hm (RADVO and OV2-SLAM) to 1.01 deg/hm (VINS-Fusion). Here, we can see that OV2-SLAM achieved the most steady and accurate location while VINS-Fusion has the lowest accuracy.

From the graph, it is shown that the performances for the feature-based technique are varied while the direct-based approach achieved more consistent results. Undeniably, both feature-based and direct-based approaches are both competitive in their performances. Interestingly, GDVO, which applied a semi-direct approach seems to be able to achieve among the lowest RTE average despite its high rotation error. This shows that their VO technique succeeded in obtaining optimum scale estimation for accurate pose estimation.

Conclusion

This study reviews and compares the VO performances according to the driving sequence environment. From the performance evaluation on the KITTI dataset, SOFT-based VO performed well in most of the sequences. It is shown that driving sequences in residential areas generally achieved good localization accuracy with an average of 0.72% for RTE. However, the localization based on VO would suffer from rotation error as incurred in one of the residential sequences where sequence 07 achieved the average of 0.48 deg/hm for its RRE due to the noise from other moving vehicles in various directions at the junction stop.

As for the VO performance on a highway road, the RTE average for all VO techniques was exceptionally high (2%) and we predict this would deteriorate as the vehicle speed increases. Since VO is targeted to facilitate vehicle positioning for better accuracy, especially during GPS signal outages and autonomous driving, we need to focus on the common condition of a positioning problem. With the growing public dataset for VO evaluation, we look forward to seeing more optimization on high-speed driving localization.

Acknowledgment

The authors would like to thank University Malaysia Sarawak for providing the facilities used in this study.

Funding Information

This study was supported by the Kementerian Pengajian Tinggi Malaysia, Fundamental Research Grant Scheme, RACER/1/2019/ICT02/UNIMAS//l.

Author's Contributions

Dayang Nur Salmi Dharmiza and Awang Salleh: Contributed to the data analysis, preparation, and development of this manuscript.

Kuryati Kipli: Helped with data interpretation, provided a critical review of the manuscript, and finalized the submission.

Ethics

This article is original and its contents are unpublished. The corresponding author confirms that the other author has read and approved the manuscript and that there are no ethical issues involved.

References

Alahi, A., Ortiz, R., & Vandergheynst, P. (2012, June). Freak: Fast retina keypoint. In *2012 IEEE conference on computer vision and pattern recognition* (pp. 510-517). IEEE. https://doi.org/10.1109/CVPR.2012.6247715

Awang Salleh, D. N. S. D., & Seignez, E. (2018). Swift path planning: Vehicle localization by visual odometry trajectory tracking and mapping. *Unmanned Systems*, *6*(04), 221-230. https://doi.org/10.1142/S2301385018500085

Bénet, P., & Guinamard, A. (2020, September). Robust and Accurate Deterministic Visual Odometry. In *Proceedings of the 33rd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2020)* (pp. 2260-2271). https://doi.org/10.33012/2020.17586.

Blanco-Claraco, J. L., Moreno-Duenas, F. A., & González-Jiménez, J. (2014). The Málaga urban dataset: High-rate stereo and LiDAR in a realistic urban scenario. *The International Journal of Robotics Research*, *33*(2), 207-214. https://doi.org/10.1177/0278364913507326.

Burri, M., Nikolic, J., Gohl, P., Schneider, T., Rehder, J., Omari, S., ... & Siegwart, R. (2016). The EuRoC micro aerial vehicle datasets. *The International Journal of Robotics Research*, *35*(10), 1157-1163. https://doi.org/10.1177/0278364915620033

Calonder, M., Lepetit, V., Strecha, C., & Fua, P. (2010, September). Brief: Binary robust independent elementary features. In *European conference on computer vision* (pp. 778-792). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-15561-1_56

Cvišić, I., & Petrović, I. (2015, September). Stereo odometry is based on careful feature selection and tracking. In *2015 European Conference on Mobile Robots (ECMR)* (pp. 1-6). IEEE. https://doi.org/10.1109/ECMR.2015.7324219.

Cvišić, I., Ćesić, J., Marković, I., & Petrović, I. (2018). SOFT-SLAM: Computationally efficient stereo visual simultaneous localization and mapping for autonomous unmanned aerial vehicles. *Journal of Field Robotics*, *35*(4), 578-595. https://doi.org/10.1002/rob.21762

Cvišić, I., Marković, I., & Petrović, I. (2022a). SOFT2: Stereo Visual Odometry for Road Vehicles Based on a Point-to-Epipolar-Line Metric. *IEEE Transactions on Robotics*. https://doi.org/10.1109/TRO.2022.3188121

Cvišić, I., Marković, I., & Petrović, I. (2022b). Enhanced calibration of camera setups for high-performance visual odometry. *Robotics and Autonomous Systems*, *155*, 104189. https://doi.org/10.1016/j.robot.2022.104189

Engel, J., Koltun, V., & Cremers, D. (2017). Direct sparse odometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(3), 611-625. https://doi.org/10.1109/TPAMI.2017.2658577

Engel, J., Schöps, T., & Cremers, D. (2014, September). LSD-SLAM: Large-scale direct monocular SLAM. In *European conference on computer vision* (pp. 834-849). Springer, Cham. https://doi.org/https://doi.org/10.1007/978-3-319-10605-2_54

Forster, C., Zhang, Z., Gassner, M., Werlberger, M., & Scaramuzza, D. (2016). SVO: Semidirect visual odometry for monocular and multicamera systems. *IEEE Transactions on Robotics*, *33*(2), 249-265. https://doi.org/10.1109/TRO.2016.2623335

Gehrig, M., Aarents, W., Gehrig, D., & Scaramuzza, D. (2021). Dsec: A stereo event camera dataset for driving scenarios. *IEEE Robotics and Automation Letters*, *6*(3), 4947-4954. https://doi.org/10.1109/LRA.2021.3068942

Geiger, A., Lenz, P., Stiller, C., & Urtasun, R. (2013). Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, *32*(11), 1231-1237. https://doi.org/10.1177/0278364913491297

Geiger, A., Ziegler, J., & Stiller, C. (2011, June). Stereoscan: Dense 3d reconstruction in real-time. In *2011 IEEE intelligent vehicles symposium (IV)* (pp. 963-968). IEEE. https://doi.org/10.1109/IVS.2011.5940405

Krešo, I., & Šegvic, S. (2015). Improving the egomotion estimation by correcting the calibration bias. In *10th International Conference on Computer Vision Theory and Applications*. https://doi.org/10.5220/0005316103470356

Leutenegger, S., Chli, M., & Siegwart, R. Y. (2011, November). BRISK: Binary robust invariant scalable key points. In *2011 International conference on Computer Vision* (pp. 2548-2555). IEEE. https://doi.org/10.1109/ICCV.2011.6126542

Liao, Y., Xie, J., & Geiger, A. (2022). KITTI-360: A novel dataset and benchmarks for urban scene understanding in 2nd and 3rd. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.48550/ARXIV.2109.13410

Maddern, W., Pascoe, G., Linegar, C., & Newman, P. (2017). 1 year, 1000 km: The Oxford RobotCar dataset. *The International Journal of Robotics Research*, *36*(1), 3-15. https://doi.org/10.1177/0278364916679498

Meilland, M., Comport, A. I., & Rives, P. (2011, August). Real-time dense visual tracking under large lighting variations. In *British Machine Vision Conference* (pp. 45-1). British Machine Vision Association. https://doi.org/10.5244/C.25.45

Mukhopadhyay, S. C. (2014). Wearable sensors for human activity monitoring: A review. *IEEE Sensors Journal*, *15*(3), 1321-1330. https://doi.org/10.1109/JSEN.2014.2370945

Mur-Artal, R., Montiel, J. M. M., & Tardos, J. D. (2015). ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, *31*(5), 1147-1163. https://doi.org/10.1109/TRO.2015.2463671

Neubeck, A., & Van Gool, L. (2006, August). Efficient non-maximum suppression. In *18th International Conference on Pattern Recognition (ICPR'06)* (Vol. 3, pp. 850-855). IEEE. https://doi.org/10.1109/ICPR.2006.479

Newcombe, R. A., Lovegrove, S. J., & Davison, A. J. (2011, November). DTAM: Dense tracking and mapping in real-time. In *2011 international conference on computer vision* (pp. 2320-2327). IEEE. https://doi.org/10.1109/ICCV.2011.6126513.

Nistér, D., Naroditsky, O., & Bergen, J. (2004, June). Visual odometry. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* (Vol. 1, pp. I-I). IEEE. https://doi.org/10.1109/CVPR.2004.1315094

Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011, November). ORB: An efficient alternative to SIFT or SURF. In *2011 International conference on computer vision* (pp. 2564-2571). IEEE. https://doi.org/10.1109/ICCV.2011.6126544

Wang, K., Huang, X., Chen, J., Cao, C., Xiong, Z., & Chen, L. (2019). Forward and backward visual fusion approach to motion estimation with high robustness and low cost. *Remote Sensing*, *11*(18), 2139. https://doi.org/10.3390/rs11182139

Wang, R., Schworer, M., & Cremers, D. (2017). Stereo DSO: Large-scale direct sparse visual odometry with stereo cameras. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3903-3911). https://doi.org/10.1109/ICCV.2017.421

Zhu, J. (2017, August). Image Gradient-based Joint Direct Visual Odometry for Stereo Camera. In *IJCAI* (pp. 4558-4564). https://doi.org/10.24963/ijcai.2017/636

Zhu, Z. A., Thakur, D., Ozaslan, T., Pfrommer, B., Kumar, V., & Daniilidis, K. (2018). The Multi Vehicle Stereo Event Camera Dataset: An Event Camera Dataset for 3D Perception. *arXiv e-prints*, arXiv-1801. https://doi.org/10.1109/LRA.2018.2800793