

Urban@CRAS dataset: Benchmarking of visual odometry and SLAM techniques

Ana Rita Gaspar*, Alexandra Nunes, Andry Maykol Pinto, Aníbal Matos

INESC TEC, Porto, Portugal

HIGHLIGHTS

- Public dataset that captures a multiplicity of conditions and urban situations.
- Dataset acquired by through a set of heterogeneous sensors (in roof of a car).
- Ground-truth with an accurate positioning information obtained from a RTK system.
- Benchmarking of vOdometry and vSLAM methods.

ARTICLE INFO

Article history:

Available online 31 August 2018

Keywords:

Dataset
Autonomous driving
Mobile robots
Odometry
SLAM
Benchmarks
Computer vision
Urban@CRAS

ABSTRACT

Public datasets are becoming extremely important for the scientific and industrial community to accelerate the development of new approaches and to guarantee identical testing conditions for comparing methods proposed by different researchers. This research presents the Urban@CRAS dataset that captures several scenarios of one iconic region at Porto Portugal. These scenarios present a multiplicity of conditions and urban situations including, vehicle-to-vehicle and vehicle-to-human interactions, cross-sides, turn-around, roundabouts and different traffic conditions. Data from these scenarios are timestamped, calibrated and acquired at 10 to 200 Hz by through a set of heterogeneous sensors installed in a roof of a car. These sensors include a 3D LIDAR, high-resolution color cameras, a high-precision IMU and a GPS navigation system. In addition, positioning information obtained from a real-time kinematic satellite navigation system (with 0.05m of error) is also included as ground-truth. Moreover, a benchmarking process for some typical methods for visual odometry and SLAM is also included in this research, where qualitative and quantitative performance indicators are used to discuss the advantages and particularities of each implementation. Thus, this research fosters new advances on the perception and navigation approaches of autonomous robots (and driving).

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

The mobile robots and self-driving cars are being considered two mega-trends for the next decade [1]. Their ability to perceive the scene is currently inadequate in many ways and so, methods with advanced capabilities to interpret and understand all static and dynamic elements of the scene are crucial to increase the autonomy level (from level 3 to 5), the thrust of users in this emerging technology and the safeness of the interaction modes [2]. Moreover, many requirements related to the autonomy level, computational power, reliability under distinct environmental conditions and sensor costs must be balanced for a fully-autonomous system [3]. In this context, the availability of data acquired under realistic (normal and hazard) circumstances contributes to the

arrival of cutting-edge methods that foster the perception capability of such systems. More diverse and truthful scenes prevent overfitting the training data during the creation of artificial intelligence (AI) entities which usually outperform existing ones [4]. Additionally, data captured from heterogeneous sensors in distinct scenarios make possible to conduct qualitative and quantitative assessments of the performance obtained by existing methods, a process often called as benchmarking.

This article presents two main contributions, that extend the research work presented in ROBOT'2017 Conference [5]: the Urban@CRAS dataset (<https://rdm.inesctec.pt/dataset/nis-2018-001>) and the benchmarking evaluation of visual SLAM (Simultaneous Localization and Mapping) and Odometry approaches for this dataset. The Urban@CRAS dataset provides a set of scenarios from a particular region of Porto that capture different interactions between static–dynamic elements with a moving car, where

* Corresponding author.

E-mail address: argaspar@inesctec.pt (A.R. Gaspar).

some sensors were installed namely, a 3D LIDAR (Light Detection And Ranging), a stereoscopic system formed by a couple of high-resolution color cameras, a multi-constellation GNSS (Global Navigation Satellite System) receiver and a high-precision IMU (Inertial Measurement Unit) navigation system. The Urban@CRAS dataset is ROS (Robot Operating System)-compliant: the data is calibrated and timestamped because sensors operate at different frequencies (10 to 200 Hz). This dataset includes diverse situations that are challenging for visual analysis namely, light changes, different scenes (urban, highway and coastal zones), dynamic elements and others.

Several vSLAM and vOdometry algorithms have been proposed by researchers for different applications. As a consequence, these algorithms have distinct behaviors under different testing scenarios. This research presents a comparative study of some visual odometry and SLAM techniques. This research presents a comparative study of some visual odometry and SLAM techniques (Direct methods are not considered in this research) that are publicly available. The performance of these methods considering application scenarios such as self-driving cars and mobile robots, are discussed. It is worthwhile to mention that, the benchmark analysis conducted in this research tries to highlight and understand the main characteristics and advantages of each implementation, considering different testing conditions. Therefore, these benchmarking results are valuable for a potential selection of method that is more convenient to a particular application.

Therefore, the main contributions of this paper include:

- A novel dataset called Urban@CRAS, provided as a public repository from CRAS (Centre for Robotics and Autonomous Systems) that is ROS-Compliant—<https://rdm.inesctec.pt/dataset/nis-2018-001>;
- Several traffic conditions are depicted in scenarios of the Urban@CRAS dataset which include vehicle-to-vehicle and vehicle-to-human interactions, cross-sides, turn-around and roundabouts. These scenarios clearly present challenges for existing vSLAM and vOdometry approaches;
- Benchmarking of typical vOdometry and vSLAM techniques. Qualitative and quantitative evaluations of several techniques are presented for different testing conditions. These experiments make possible to highlight the major properties, advantages and disadvantages of each implementation. Accurate positioning data is provided for the majority of scenarios from the Urban@CRAS dataset – a ground-truth – where information is obtained from a dual-frequency real-time kinematic (RTK) satellite navigation system with a Euclidean error lower than 0.05 m.

This paper is organized as follows: Section 2 presents methods for vOdometry and vSLAM, and describes some datasets that are available to the scientific and industrial community. The Urban@CRAS dataset is described in Section 3. Afterwards, Section 4 presents a comparative study where results from several vOdometry and vSLAM techniques are discussed for both traditional and Urban@CRAS datasets. Finally, the major conclusions of this work are discussed in Section 5.

2. Related works

Nowadays, datasets are relevant for researchers to validate techniques and implementations such as, images analysis and sensor fusing. In addition, they are relevant for benchmarking of different techniques and to support the selection of a suitable method for a specific application [6]. Over the past few years, several datasets have been proposed by research teams, for instance, EUROC [7], TUM [8], MIT Stata Center [9], KITII [10] and New College [11]. As in this research the proposed focus is the autonomous driving,

to obtain more similar conditions to intended environment, the outdoor datasets will be more carefully considered.

The KITTI¹ dataset includes 22 stereo images sequences of different trajectories in urban and highway environments. The dataset is composed by high resolution image sequences, LIDAR and GPS data for all trajectories. The MIT Stata Center² represents the indoor environment captured by a mobile robot during long paths. This dataset captures several directions changes and loop-closures which makes this dataset suitable for SLAM techniques. The New College³ dataset provides images, range and intensity and location data (GPS and IMU). This dataset was acquired in gardens. In Table 1 is presented an overview of the datasets of the literature, based on environment, data provided, availability (✓) of the ground-truth and some relevant details.

The datasets are quite relevant to study the limitation of distinct methods that estimate the egomotion: vOdometry and vSLAM. Currently, non-structured and dynamic environments usually inflict severe challenges for visual-based techniques because conventional implementations of vOdometry and vSLAM implementations do not detect nor ignore the presence of dynamic elements that are in the scenario. The literature about vOdometry and vSLAM is large and so, it is difficult to assess what method has a suitable behavior when subjected to different scene conditions. A set of methods were selected from the scientific community by taking into consideration the availability of the code and the relevance of the method.

First, vOdometry methods such as, mono-vo [12], viso2 [13] and mORB-SLAM [14] are being applied to mobile robots [15–17]. The mono-vo [12] implementation developed in 2015 uses Features from Accelerated Segment Test (FAST) and the Kanade–Lucas–Tomasi approach to search correspondences between the current and the next image. It includes a mechanism to search for new features and removes outliers. Moreover, this implementation uses information from an external source of data to remove the ambiguity of scale. Lastly, the mono-vo follows a heuristic to estimate the dominant motion. The viso2 [13] implementation was developed in 2011 to estimate the camera position through a set of rectified images. The method resorts Oriented FAST and Rotated BRIEF (ORB) features. The technique is flexible because it has a large number of configurable parameters, but could be difficult to set up when compared to mono-vo. The viso2 assumes that the camera motion is at fixed height from the ground (used to predict a scale factor). Besides that, this method uses a bucketing technique to correct distribution of features however, a pure rotation affects the estimation process. Therefore, the method estimates the egomotion during movements with 6 Degrees of Freedom (6DoF). The mORB-SLAM is a monocular implementation of the ORB-SLAM2, presented in [14]. A keyframe approach is used to estimate the egomotion. The ORB features are selected through a bucketing strategy followed by a procedure that removes outliers. At the end, a bundle adjustment operation uses the keyframes to provide the displacement of the observer. A bundle adjustment is conducted between keyframes.

The vSLAM techniques are helpful for the active exploration of unknown environments [18–20]. Implementing a vSLAM approach is usually more complex than vOdometry because the formulation must include the detection of loop-closures (revisited areas). To decrease the positioning error, the navigation strategy of a mobile robot with a vSLAM is modified to increase the number of loop-closures across the path. Thus, Common vSLAM techniques in the scientific literature include, the RTAB-Map [21], S-PTAM [22] and ORB-SLAM2 [14] implementations. The RTAB-Map [21] technique

¹ http://www.cvlibs.net/datasets/kitti/eval_odometry.php.

² <http://projects.csail.mit.edu/stata/downloads.php>.

³ <http://www.robots.ox.ac.uk/NewCollegeData/>.

Table 1
Overview of state-of-art datasets.

Dataset	KITTI	MIT stata center	New college
Environment	Outdoor (urban and highway)	Indoor	Outdoor (gardens)
Data	- Stereo images - 3D points - Location (GPS)	- Stereo images - Laser	- Stereo images - Location (GPS and IMU)
Ground-Truth	✓	✓	✓
Details	- Long sequences - Loop-closure	- Long paths - Loop-closure - Dark areas (less features) - Several direction changes	- Loop-closure - Dark areas (less features)



Fig. 1. Scenarios represented in the Urban@CRAS dataset (real road situations during several trajectories).

was introduced in 2014, for scenarios involving cars. The method follows a graph-based approach with an incremental loop-closure detection. The current implementation of RTAB-MAP makes possible to use several vOdometry methods (e.g., viso2) to represent the movement of the observer. The loop closure detection is constructed online through a bag-of-words approach and Speeded-Up Robust Features (SURF) descriptors. The RTAB-Map implements a memory management approach to consider only a part of the map for online processing. The S-PTAM [22] was presented in 2015. This implementation divides the SLAM approach into two separated operations: tracking and map optimization. The method resorts to a BRIEF descriptor to reduce the storage requirement and to speed up the feature correspondence procedure. Moreover, a Shi-Tomasi algorithm with a bucketing technique enhances the distribution of features that were previously detected. Finally, this method presents an iterative bundle adjustment and no loop-closure strategy. The ORB-SLAM2 [14] implementation appeared in 2015 and can be used for monocular, stereo and RGB-D cameras. The process that estimates the egomotion is more robust to different viewpoints and illumination changes. Keyframes with a bag-of-words approach are defined for motion estimation and relocalization (loop-closure) of the observer in the scenario. In fact, a covisibility graph increases the efficiency of loop-closure because the similarity is calculated only for neighbors of a graph representation (and strong matches will be used to update an essential graph) [23].

3. Urban@CRAS dataset

The Urban@CRAS dataset comprises a set of trajectories conducted across different locations between Porto and Matosinhos, in Portugal. It provides sensor information obtained mainly by high resolution cameras, LIDAR, GPS and IMU that were installed on the roof of a car from INESC TEC. This dataset was acquired in a sunny day, provides realistic insights of urban road environments and captures dynamic interactions with many road elements (such as, moving cars, motorcycles and pedestrians), light changes, traffic conditions and heterogeneous scenarios: a coastal zone, avenues, roundabouts and highways. These features are valuable for the development of systems for situation awareness, for the navigation of mobile robots and autonomous driving (up to level 5). Moreover, it can be used for benchmarking both vOdometry and vSLAM techniques (since there are trajectories with loop-closures incorporated by the trajectories). Fig. 1 illustrates some scenarios captured in Urban@CRAS dataset. Fig. 2 shows an example of the three-dimensional representation of the scene obtained by the LIDAR. This data is available for all the scenarios included in the Urban@CRAS.

The Urban@CRAS is ROS-compliant and sensor data collected from scenarios is stored in bag files. The ROS is well adopted (e.g., in the field of mobile robotics) since it preserves the output rate of sensors, data that is timestamped and ready to use.

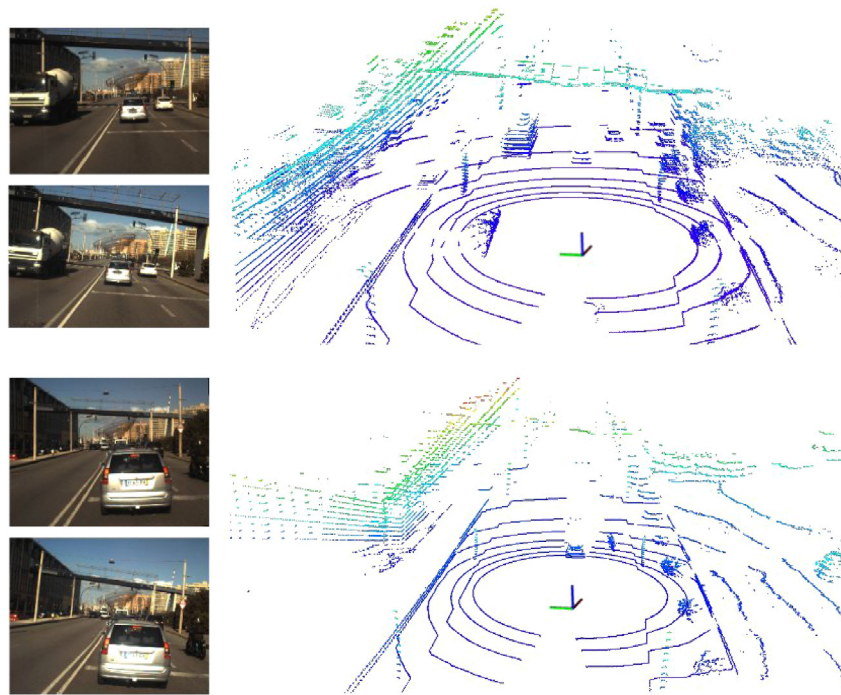


Fig. 2. The 3D representation of two scenes based on data from the LIDAR sensor (the images obtained by the stereoscopic system are also presented on the left).

3.1. Scenarios

A set of 5 scenarios is included in the Urban@CRAS dataset which tries to represent some frequent situations and events of the road transport mode. These scenarios present several challenges for road perception solutions of scene understanding based on the current perception solutions (e.g., vSLAM and vOdometry) for mobile robots or autonomous driving: (1) vehicle-to-vehicle and vehicle-to-human interactions (moving cars, trucks, motorcycles and pedestrians), (2) driving actions (cross-sides, turn-around and roundabouts), (3) different road/environment conditions, and (4) incorrect driver behavior.

The scenarios presented in Urban@CRAS dataset are the following:

- *DLoop* - part of the “Boavista” Avenue was covered in both directions. The trajectory includes two complete rounds across a roundabout near to Fort São Francisco (“Castelo do Queijo”). The scenario has loop-closures, crossroads, vehicles moving in the opposite direction, turn-around locations and pedestrians on the road (2.4 km). Moreover, there are illumination changes and several dark areas;
- *CoastalLoop* - the trajectory starts in the “Boavista” Avenue and follows a coastal road that connects two roundabouts namely, the “Fort São Francisco” to the “Anémoma”. The “Anémoma” roundabout called “Anémoma” is circumvented and the path ends close to the initial position (1.8 km). The scenario has loop-closures, crossroads, traffic lights, vehicles moving in the opposite direction, turn-around locations and light changes (bright and dark areas);
- *UrbanLoop* - the path extends the movement of the car on the “Boavista” Avenue (in both directions) and includes a single pass in the roundabout near to Fort São Francisco. The path ends in the initial position (2 km). The scenario has loop-closures, crossroads, traffic lights, vehicles moving in the opposite direction, turn-around locations, trees, pedestrians in the road and a high number of buildings;

- *Avenue* - the path covers a large part of the “Boavista” Avenue up to one highway access (3 km). This scenario does not have loop-closure but presents similar conditions as the UrbanLoop in a linear movement (without roundabouts or a turn-around location);
- *Highway* - covers a highway between the “Boavista” Avenue and INESC TEC headquarters (5 km). The scenario does not have a loop-closure and present significant light changes (bright and dark areas), many vehicles in fast motion, traffic signs lights and many overpasses.

3.2. Sensor setup

Fig. 3 depicts the sensor setup that was installed in a car to acquire data. Coordinate frames that provide the physical arrangement of sensors are also presented in this figure.

As can be noticed, this setup includes different sensors organized in a cost-effective solution that comprises⁴:

- 1 x MTi-30 Xsens IMU, 200 Hz, 0.2°/0.5° angular accuracy;
- 1 x Swift Navigation GPS receiver, 20 Hz, L1/L2 RTK, 0.01 m horizontal accuracy and 0.015 m vertical accuracy;
- 2 x Mako G-125 with a 6 mm lens at 30 frames per second (fps) and a resolution of 1292 × 964;
- 1 x Velodyne VLP-16, 4 Hz, 0.03 m of range accuracy, measuring distance up to 100 m and field of view of 360° horizontal and 30° vertical.

The GPS receiver provides location data in terms of latitude and longitude but also incorporates the RTK technology which is able to achieve a centimeter-level accuracy. A processing module (Intel i7-4700HQ CPU @ 2.40 GHz) for data acquisition receives up-to-date sensor information using ethernet communication.

⁴ The information of sensors were obtained from the datasheet of vendors.

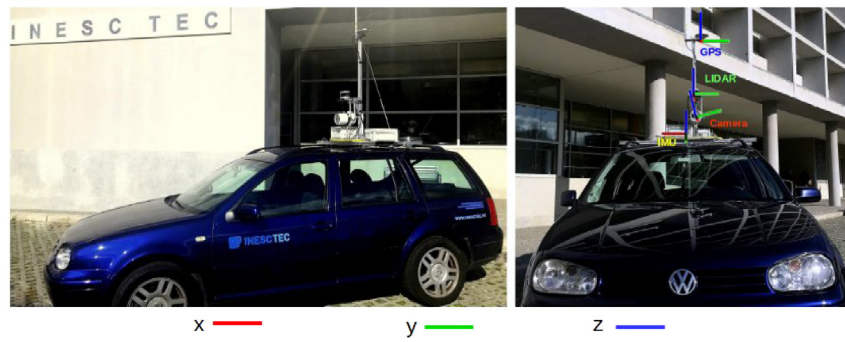


Fig. 3. The setup of sensors used in Urban@CRAS dataset. The car (on the left) and the representation of the coordinate systems (on the right).

3.3. Sensor calibration and synchronization

Extrinsic calibrations are provided in the dataset since they represent the physical displacements and orientations of each sensor to a reference coordinate system. The stereoscopic system formed by two cameras was calibrated according to [24], that uses a chessboard and calculates the extrinsic parameters from corner extraction. The intrinsic parameters of the LIDAR are provided by the vendor.⁵ The relation between coordinate frames of the stereo-rig and the LIDAR were obtained by the single value decomposition of a set of 3D correspondences [25]. Finally, the IMU and the GPS frames were manually measured in relation to a reference coordinate system.

The synchronization of sensors can be verified by the timestamp information of the ROS messages. This timestamp means the moment that data is acquired by the processing system regarding heterogeneous sensors with different output rates.

3.4. Data description

The information provided by the sensors used by the Urban@CRAS dataset can be found in the following ROS topics (see Table 2). The trajectory made by the car is also available for some scenarios. This trajectory has a centimeter-level accuracy (obtained by RTK) and can be considered as the ground-truth by many applications which can be helpful for benchmarking. The RTK technique estimates the pose of the car to a base station by using GPS locations and communication channels. Therefore, this ground-truth makes it possible to estimate the Euclidean error of techniques that provide navigation information based on LIDAR or visual sensors (e.g., vOdometry and vSLAM). This ground-truth is available for the DLoop (complete) and CoastalLoop (partial) scenarios of Urban@CRAS, see Table 3 and Fig. 4. Only a few datasets that are currently available on the web contain this type of information.

4. Results of benchmarking

A comparative evaluation of vOdometry and vSLAM techniques of Section 2 is demonstrated in this section. A set of experiments was conducted using publicly available datasets and the URB@CRAS dataset. The first experiment aims to provide a comparative analysis of the behavior of vOdometry and vSLAM methods for traditional datasets, see Section 4.1. Afterwards, Section 4.2 evaluates the second experiment aimed to evaluate a technique (vOdometry and vSLAM) that obtained a good performance in the previous experiment for a challenging scenario on Urban@CRAS.

Implementations of vOdometry and vSLAM methods used in this research are public, changes in their parameters were kept at

Table 2

List of ROS topics provided.

Sensor	ROS topic	Summary
IMU (200 Hz)	/imu/data	- Angular velocity (x,y,z,cov)
		- Linear acceleration (x,y,z,cov)
GPS (20 Hz)	/gps/fix	- Orientation (w,x,y,z)
		- Latitude
LIDAR (10 Hz)	/velodyne_points	- Longitude
		- Altitude
Camera (15 Hz)	/stereo/right/camera_info	- Covariance measurements
	/stereo/left/camera_info	- 3D points
	/stereo/right/image_raw/compressed	- Calibration
	/stereo/left/image_raw/compressed	- Raw images

Table 3

Ground-truth availability for all scenarios.

Trial	Ground-truth
DLloop	Yes
CoastalLoop	Partial
UrbanLoop	Partial
Avenue	Partial
Highway	No

a minimum-level, and the results were obtained by a i7-6700 @ 2.6 GHz × 8 processor computer without graphics processing unit (GPU).

4.1. Comparison using other traditional datasets

This section compares the performance of distinct vOdometry and vSLAM approaches. These techniques were evaluated in the KITTI⁶ and the New College⁷ datasets to analyze the behavior of visual-based approaches for traditional scenarios (discussed in Section 2). Performance indicators such as, the number of frames processed in each sequence, the processing time and the (normalized) Euclidean error are considered in the benchmarking.

The monocular implementations of vOdometry approaches are evaluated first. Paths obtained without loop-closures are normalized to compensate the unknown scale factor. This methodology makes it possible to compare the accuracy of the egomotion estimation that were obtained by each method. The vSLAM methods

⁵ VeloView-VLP-16-HiRes.xml on December 29, 2016.

⁶ http://www.cvlibs.net/datasets/kitti/eval_odometry.php.

⁷ <http://www.robots.ox.ac.uk/NewCollegeData/>.



Fig. 4. Visual representation of the ground-truth for the DLoop scenario (on the left) and CoastalLoop scenario (on the right).

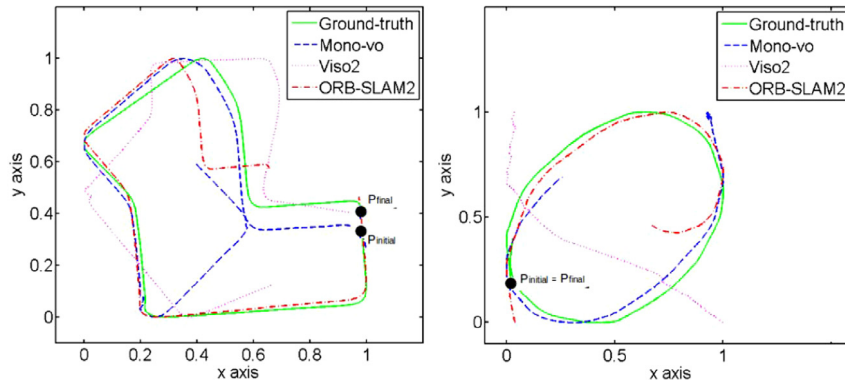


Fig. 5. The normalized trajectories obtained (with ground-truth) by vOdometry techniques with the sequence 07 of KITTI dataset (on the left) and with the New College (on the right) dataset.

are evaluated in terms of number of loop-closures detected and accuracy. The accuracy is discussed by selecting two checkpoints in each scenario (“p1” and “p2”). Finally, this research discusses the aliasing phenomenon for the vSLAM approaches, which is usually ignored in the literature.

Techniques for vOdometry were evaluated for the KITTI (sequence 07) and New College datasets. Fig. 5 depicts the trajectory made by mono-vo [12], viso2 [13] and mORB-SLAM [14].

Fig. 5 depicts that the mono-vo implementation follows the movement of the camera for most of the time however, an incorrect direction change causes a high positioning error at the end of the trajectory. On the other hand, the mORB-SLAM implementation estimates the camera pose for the sequence since direction changes were captured with high accuracy. The viso2 had some difficulties in capturing the direction changes due to some oscillations of the camera. As a consequence, the method returned a positioning error that is high.

Table 4 shows the results of several performance indicators that were collected by vOdometry techniques in both scenarios (sequence 07 of KITTI and New College).

As discussed previously, the mORB-SLAM estimated the egomotion of the observer with high accuracy since the (normalized) average error was less than 0.1 and 0.4 for KITTI and New College scenarios, respectively. Although the maximum and average errors of the path resulted by mORB-SLAM were (almost) half of the other techniques, it took more time to process both sequences. The mono-vo is the technique with higher maximum error in both trajectories. The method captures the movement of the observer using consecutive frames, reason why the resulting trajectory has a profile that is similar to the ground-truth. Finally, the viso2 implementation originated a final trajectory with positioning errors of 0.32 and 0.85 for the (sequence 07) KITTI and New College scenarios, respectively. Although the viso2 was able to detect changes of direction in the first scenario, it was unable to represent

the movement performed in the second scenario. The method is affected by shadows, vibrations of the camera and relative changes in depth of the view. All these factors contribute to the behavior of viso2 however, the technique is computationally efficient since it took less time to compute all frames.

Techniques for vSLAM were evaluated for the KITTI dataset: 05 and 09 sequences. Fig. 6 depicts the trajectory made by RTAB-Map [21], ORB-SLAM2 [14] and S-PTAM [22] for both sequences.

The ORB-SLAM2 was able to detect all loop-closures for the first sequence and thus, the trajectory was adjusted properly. The technique did not detect any revisited area in the second scenario and, consequently, the trajectory was not adjusted by the global bundle adjustment. In fact, the sequence 05 of KITTI was challenging for the majority of methods but the ORB-SLAM2 was the only implementation to follow correctly the trajectory made by the car. This method achieved the best performance in terms of positioning error (0.55 and 1.05 m for both checkpoints in the first scenario and 11.46 and 37.62 m for the second scenario). However, the method presented a processing time that is higher when compared to other techniques.

Table 5 demonstrates the reduced performance of RTAB-Map and S-PTAM for the sequence 05 of KITTI. All methods tried to replicate the movement of the observer however, there are deviations from the original route that could be caused by vibrations and dark areas in images, affecting the bucketing process. As a consequence, the S-PTAM was only able to estimate the egomotion for half of the trajectory from the second scenario, causing the aliasing phenomenon and high positioning errors (60.61 and 103.25m for both checkpoints). Finally, the RTAB-Map did not estimated a realistic path for both scenarios. A potential reason for this issue can be relate to the aliasing phenomenon and no detection of loop-closure.

Table 4

The normalized trajectories obtained by KITTI (07) and New College datasets. KFr represents the number of keyframes used for egomotion estimation. The performance metrics are included namely, the number of frames processed by the method, the maximum and average error normalized for the entire trajectory and the processing time for the scenario.

	Methods	Processed frames	Normalized error			Time
			max.	avg.	std	
KITTI (07)	Mono-vo	1000	0.64	0.13	0.17	85 s
	Viso2	999	0.62	0.32	0.18	74 s
	mORB-SLAM	996	0.36	0.10	0.10	109 s
		KFr = 374				
New College	Mono-vo	2500	1.22	0.61	0.26	127 s
	Viso2	1765	1.11	0.85	0.19	124 s
	mORB-SLAM	1657	0.69	0.37	0.16	273 s
		KFr = 293				

Table 5

The trajectories obtained by vSLAM methods in the KITTI dataset (05 and 09 sequences). The performance metrics are included namely, the number of processed frames by each method, the maximum error for the entire trajectory, the processing time for the all scenario and the average error for checkpoints (p1 and p2).

	Methods	Processed frames	Error		Time
			max.	avg.: p1 p2	
KITTI (05)	ORB-SLAM2	2761	1.64 m	0.55 m 1.05 m	5 m32 s
	RTAB-Map (viso2)	1224	420.91 m	30.84 m 17.73 m	4 m57 s
	S-PTAM	1916	123.29 m	33.05 m 8.41 m	4 m43 s
KITTI (09)	ORB-SLAM2	1591	28.87 m	11.46 m 37.62 m	3 m4 s
	RTAB-Map (viso2)	648	145.14 m	— 207.65 m	2 m49 s
	S-PTAM	884	115.06 m	60.61 m 103.25 m	2 m40 s

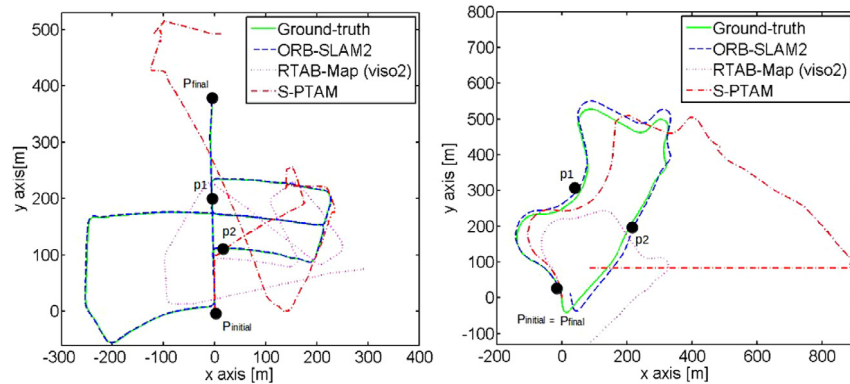


Fig. 6. The trajectories obtained by vSLAM techniques for the sequences 05 (on the left) and 09 (on the right) of KITTI dataset (with ground-truth).

4.2. Comparison using Urban@CRAS dataset

The performance of a some vOdometry and vSLAM methods is analyzed in this section namely, viso2 and ORB-SLAM2. Results obtained during the benchmarking of the previous section make possible to select only these two techniques since the ORB-SLAM2 indicated to be the best vSLAM implementation in terms of accuracy of the trajectory and reliability for detecting loop-closures; and the viso2 was computational more efficient than other vOdometry methods and captured the trajectory profile with some level of accuracy. Therefore, the viso2 and ORB-SLAM2 techniques are validated in the Urban@CRAS dataset, in particular, the Coastal-Loop and DLoop scenarios respectively. These scenarios present challenges such as, dynamic elements in the road (moving cars, trucks and motorcycles) and lead to more realistic scenes that are crucial for evaluating the performance of visual-based methods for autonomous driving applications and mobile robots.

The viso2 was evaluated in the CoastalLoop scenario. This scenario presents an observer moving between two roundabouts which means several changes in direction. The analysis started in the roundabout near to Fort São Francisco. The detection of loop-closures was avoided in this evaluation – vOdometry techniques – and so, the scenario was not completed. Fig. 7 demonstrates the trajectory obtained by the viso2 as well as the ground-truth (from a RTK with a centimeter-level accuracy).

As can be noticed, changes in direction and the overall movement of the car was captured, in particular, for the path that connects the two roundabouts (Fort São Francisco and Anémona). The overall accuracy of the path generated by this technique is reduced since the track of motion was lost for the beginning of the trajectory and during the way back to the first roundabout (on the right). The quality of the path generated by viso2 was severely affected by the presence of dynamic elements in the scene since features from both static and dynamic objects were extracted and

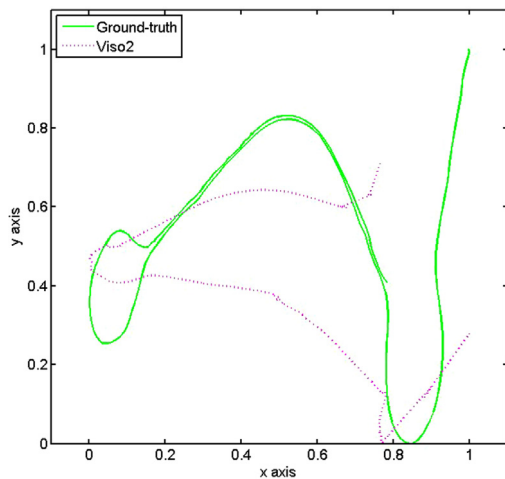


Fig. 7. The normalized trajectory of viso2 obtained in the CoastalLoop scenario of Urban@CRAS dataset (with ground-truth).

used for estimating the egomotion, causing a clear violation of the motion coherence assumption. Moreover, the scene depicted in the CoastalLoop scenario has different light conditions (related to the position of the sun in the sky) which combined with a rotational motion of the car (in roundabouts), it imposes real challenge for visual perception methods. These issues affect the accuracy of the path created by viso2.

The ORB-SLAM2 was evaluated in the DLoop scenario that captures several loop-closures and direction changes: the roundabout of Fort São Francisco and turn-arounds in “Boavista” Avenue. This scenario is suitable for evaluating the performance of vSLAM under realistic road conditions.

The path generated by the ORB-SLAM2 is presented in Fig. 8 (on the left). The ORB-SLAM2 started this scenario with an offset up to 200 m and a estimation of the egomotion that captures the first turn-around (after 700 m). The presence crossroads and vehicles moving in opposite and transversal directions can justify these issues, as well as, the overall performance of the method for the entire scenario. After the first turn-around, the ORB-SLAM2 method was able to estimate the trajectory made by the car with an error less than 100 m (caused by vehicle-to-vehicle interactions that appear in field-of-view of the stereoscopic cameras). Although the movement of the observer in the roundabout was successfully captured, the coastal perspective depicted in this scenario reduces

the number of features used in the bundle adjustment. The ORB-SLAM2 detected all loop-closures however, the final position of the trajectory was not achieved because of a darker area at roundabout that causes a bad performance for tracking features. This behavior is clearly depicted in Fig. 8 (on the right) and demonstrates that even sophisticated perception methods from the state-of-the-art have difficulties when subjected to realistic urban road environments.

A significant work has to be made by researchers to support the development of autonomous driving and mobile robots. Therefore, the Urban@CRAS dataset and the benchmarking proposed in this research work provide the basis for improving the methods of vOdometry and vSLAM.

5. Conclusion

A benchmarking of visual Odometry and SLAM techniques and a novel dataset called Urban@CRAS is presented in this research. The Urban@CRAS dataset captures realistic traffic conditions that comprise vehicle-to-vehicle and vehicle-to-human interactions, cross-sides, turn-around and roundabouts. It provides sensor information from a LIDAR, GPS, IMU and a stereoscopic camera, acquired by a car moving in several urban and highway scenarios close to an one iconic region of Porto, in Portugal.

An extensive set of experiments were conducted as part of this work to evaluate the performance of vOdometry (mono-vo [12], viso2 [13] and mORB-SLAM [14]) and vSLAM (RTAB-Map [21], S-PTAM [22] and ORB-SLAM2 [14]) techniques when subjected several scenarios. Overall, the majority of these methods performed well under traditional scenarios (one-way roads without dynamic elements in the scene) however, this research demonstrated that the performance of current vOdometry and vSLAM implementations is significantly reduced for more challenging scenarios (presented in Urban@CRAS dataset). For instance, the viso2 is affected by dynamic elements in the scene and rotational motion; and the ORB-SLAM2 estimated the egomotion with lower errors (less than 80% when compared to other vSLAM methods) for traditional scenarios and 500 m for roads having vehicles moving in opposite direction.

The vOdometry and vSLAM implementations are more and more relevant for autonomous driving and mobile robots. In this context, this research proposes the Urban@CRAS dataset that captures new perspectives and challenges cannot be found in other datasets, representing a valuable contribution to both scientific and industrial community.

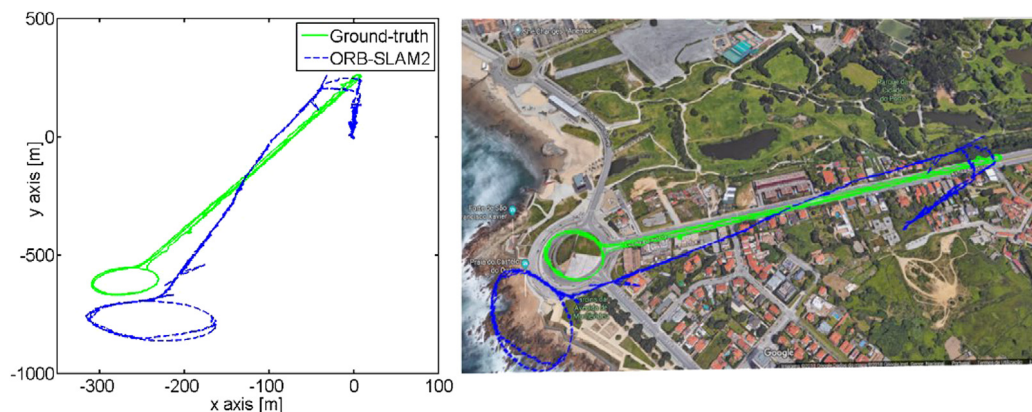


Fig. 8. Representation of the path obtained by ORB-SLAM2 for the DLoop scenario (blue line), with a ground-truth (green line) of the trajectory made by the car. Graphical representation of the path is demonstrated on the left image and a satellite view is provided on the right. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Acknowledgments

The authors would like to acknowledge Yulia Karimova by her contribution for the setup of the repository of Urban@CRAS dataset.

This work is financed by the ERDF—European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation—COMPETE 2020 Programme within project “POCI-01-0145-FEDER-006961”, and by National Funds through the FCT—Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) as part of project UID/EEA/50014/2013.

References

- [1] A. Carnevale, Will robots know us better than we know ourselves? *J. Rob. Auton. Syst.* (86) (2016) 144–151, <http://dx.doi.org/10.1016/j.robot.2016.08.027>.
- [2] N. Kalra, S.M. Paddock, Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *J. Transp. Res. Part A* (94) (2016) 182–193, <http://dx.doi.org/10.1016/j.tra.2016.09.010>.
- [3] A. Pinto, P. Costa, M. Correia, A. Matos, A. Moreira, Visual motion perception for mobile robots through dense optical flow fields, *J. Rob. Auton. Syst.* (87) (2017) 1–14, <https://doi.org/10.1016/j.rcim.2012.06.002>.
- [4] A. Pinto, L. Rocha, Correia, A. Moreira, Object recognition using laser range finder and machine learning techniques, *J. Rob. Comput. Integr. Manuf.* (29) (2013) 12–22, <http://dx.doi.org/10.1016/j.rcim.2012.06.002>.
- [5] A.R. Gaspar, A. Nunes, A. Pinto, A. Matos, Comparative Study of Visual Odometry and SLAM Techniques. ROBOT 2017: Third Iberian Robotics Conference. (694), 2018, pp. 463–474, http://dx.doi.org/10.1007/978-3-319-70836-2_38.
- [6] W. Maddern, G. Pascoe, C. Linegar, P. Newman, 1 Year, 1000Km: The oxford robotcar dataset, *Int. J. Robot. Res.* (36) (2017) 3–15, <http://dx.doi.org/10.1177/02783649166779498>.
- [7] M. Burri, J. Nikolic, P. Gohl, T. Sc1heneider, J. Rehder, S. Omari, M. Achtelik, R. Siegwart, The EuRoC mav datasets, *Int. J. Robot. Res.* (35) (2016) 1157–1163, <http://dx.doi.org/10.1177/0278364915620033>.
- [8] J. Sturm, N. Engelhard, F. Endres, W. Burgard, D. Cremers, A benchmark for the evaluation of RGB-D SLAM systems, in: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 573–580, <http://dx.doi.org/10.1109/IROS.2012.6385773>.
- [9] M. Fallon, H. Johannsson, M. Kaess, J. Leonard, The MIT stata center dataset, *Int. J. Robot. Res.* (32) (2013) 1695–1699, <http://dx.doi.org/10.1177/0278364913509035>.
- [10] A. Geiger, P. Lenz, C. Stiller, R. Urtasun, Vision meets Robotics: The KITTI Dataset, *Int. J. Robot. Res.* (32) (2013) 1231–1237, <http://dx.doi.org/10.1177/0278364913491297>.
- [11] M. Smith, I. Baldwin, W. Churchill, R. Paul, P. Newton, The new college vision and laser data set, *Int. J. Robot. Res.* (28) (2009) 595–599, <http://dx.doi.org/10.1177/0278364909103911>.
- [12] A. Singh, An OpenCV based implementation of Monocular Visual Odometry, in: *Indian Institute of Technology Kanpur. Technical report*, Kanpur, 2015.
- [13] B. Kitt, A. Geiger, H. Latgahn, Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme, in: *IEEE Intelligent Vehicles Symposium*, University of California, San Diego, CA, USA, 2010, pp. 486–492, <http://dx.doi.org/10.1109/IVS.2010.5548123>.
- [14] R. Mur-Artal, J.M.M. Montiel, J.D. Tardós, ORB-SLAM: a versatile and accurate monocular SLAM system, *IEEE Trans. Robot.* 31 (5) (2015) 1147–1163, <http://dx.doi.org/10.1109/TRO.2015.2463671>.
- [15] A. Escalera, E. Izquierdo, D. Martín, B. Musleh, F. García, J. Armingol, Stereo visual odometry in urban environments based on detecting ground features, *J. Rob. Auton. Syst.* 80 (2016) 1–10, <http://dx.doi.org/10.1016/j.robot.2016.03.004>.
- [16] L. An, X. Zhang, H. Gao, Y. Liu, Semantic segmentation-aided visual odometry for urban autonomous driving, *Int. J. Adv. Rob. Syst.* (2017) 1–11, <http://dx.doi.org/10.1177/1729881417735667>.
- [17] S. Bag, V. Venkatachalapathy, R. Ptucha, Motion estimation using visual odometry and deep learning localization, in: *Autonomous Vehicles and Machines on International Symposium on Electronic Imaging*, 2017, pp. 62–69, <http://dx.doi.org/10.2352/ISSN.2470-1173.2017.19.AVM-022>.
- [18] M. Milford, G. Wyeth, Hybrid robot control and SLAM for persistent navigation and mapping, *J. Rob. Auton. Syst.* (58) (2010) 1096–1104, <http://dx.doi.org/10.1016/j.robot.2010.05.004>.
- [19] J. Clemens, T. Reineking, T. Kluth, An evidential approach to SLAM, path planning, and active exploration, *Internat. J. Approx. Reason.* (73) (2016) 1–26, <http://dx.doi.org/10.1016/j.ijar.2016.02.003>.
- [20] F. Cheein, N. Lopez, C. Soria, F. Sciascio, F. Pereira, R. Carelli, SLAM algorithm applied to robotics assistance for navigation in unknown environments, *J. Neuroeng. Rehabilitation* (7) (2010) 1–16, <http://dx.doi.org/10.1186/1743-0003-7-10>.
- [21] M. Labbé, F. Michaud, Online global loop closure detection for large-scale multi-session graph-based SLAM, in: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, <http://dx.doi.org/10.1109/IROS.2014.6942926>.
- [22] T. Pire, T. Fischer, J. Civera, P. Cristóforis, J.J. Berles, Stereo parallel tracking and mapping for robot localization, *Intell. Rob. Syst.* (2015) 1373–1378, <http://dx.doi.org/10.1109/IROS.2015.7353546>.
- [23] D. Galvez-Lopez, J.D. Tardós, Bags of binary words for fast place recognition in images sequences, *Intell. Rob. Syst.* 28 (5) (2012) 1188–1197, <http://dx.doi.org/10.1109/IROS.2012.2197158>.
- [24] Y.M. Wang, Y. Li, J.B. Zheng, A camera calibration technique based on opencv, in: *International Conference on Information Sciences and Interaction Sciences*, 2010, pp. 403–406, <http://dx.doi.org/10.1109/ICIS.2010.5534797>.
- [25] A. Dhall, K. Chelani, V. Radhakrishnan, K.M. Krishna, LiDAR-Camera Calibration using 3D-3D Point correspondences, in: *Cornell University Library*, 2017.



Ana Rita Gaspar I finished the Master Course in Electrical and Computer Engineering at the Faculty of Engineering of the University of Porto in February 2017. I started the collaboration with the Centre for Robotics and Autonomous Systems at INESC TEC when I was developing my dissertation. This work aimed at the development of a visual navigation and mapping simultaneous system in close range situations, with the development of a visual vocabulary online method to allow the autonomous underwater vehicles to recognize the revisited areas. Now, since May 2017 I am a researcher at the Centre and I was involved in the project with a localization system based on GPS receivers and an inertial system. At this moment, I work on the perception and vision areas.



Alexandra Nunes In February 2017, I finished the Master Course in Electrical and Computer Engineering at the Faculty of Engineering of the University of Porto. The connection with the Centre for Robotics and Autonomous Systems at INESC TEC was initiated with the development of the my master thesis, which had main goal the mapping of the sea-floor as well as of the its underwater structures, using for that a visual method for egomotion estimation. Since May 2017, I am a researcher at the Centre. I have been involved in the development of a localization system based on GPS receivers and an inertial system and now, I work on the perception and vision areas.



Andry Pinto I am a senior researcher at Centre for Robotics and Autonomous Systems (CRAS) at INESC TEC. I have received a MSc. degree in Electrical and Computer Engineering from the Faculty of Engineering of the University of Porto (FEUP) in 2005, and a Ph.D. from the Department of Electrical and Computer Engineering of FEUP, Portugal, in 2014. Currently, my research interests include artificial intelligence, robotics, visual motion perception, motion analysis, optical flow, unsupervised segmentation, 3D reconstructions and underwater imaging.



Aníbal Matos received a PhD in Electrical and Computer Engineering from Porto University in 2001. He is currently coordinator of the Centre for Robotics and Autonomous Systems at INESC TEC and also an assistant professor at the Faculty of Engineering of Porto University. His main research interests are related to perception, sensing, navigation, and control of autonomous marine robots, being the author or co-author of more than 80 publications in international journals and conferences. He has participated and lead several research projects on marine robotics and its application to monitoring, inspection, search and rescue, and defense.